

FILE COPY

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

AM protein - protein search, using SW model

Run on: February 28, 2003, 16:04:04 : Search time 50 Seconds
(without alignments)
3800.308 Million cell updates/sec

Title: US-09-915-543-15

Perfect score: 7692

Sequence: 1 MSHSNPKVKSFSFGNTQSSP.....ADVGMGFSQGPENPMNF 1426

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched:

908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters:

908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Fried. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7457.5	97.0	1435	22	ABR11808
2	2026.5	26.3	1494	23	AAU78460
3	1228.5	16.0	738	23	AAU78463
4	610.5	7.9	1685	22	ABG04839
5	610.5	7.9	1693	22	ABG15619
6	578	7.5	1196	13	AAK28916
7	571	7.4	1669	23	ABR57334
8	565	7.3	1064	17	AAK93254
9	565	7.3	1064	19	AAK57652
10	565	7.3	1065	14	AAK37741

11	559.5	7.3	1047	22	ABR70874
12	558.5	7.3	2703	22	ABR63299
13	555.5	7.2	1838	15	ABR53257
14	553	7.2	1690	22	AAK23916
15	543.5	7.1	1466	22	ABR50291
16	543.5	7.1	1466	22	ABR90747
17	543.5	7.1	1469	22	ABG15191
18	536	7.0	1669	22	AAK40863
19	535	7.0	1669	23	ABR90760
20	533.5	6.9	1672	22	AAK39077
21	528	6.9	1078	21	AAV96125
22	528	6.9	1078	23	ABR80736
23	528	6.9	1078	23	ABR09628
24	528	6.9	1078	23	AAE16478
25	527	6.9	1078	16	AAK71704
26	524	6.8	1466	22	AAE02537
27	519.5	6.8	1963	22	ABR62819
28	519	6.7	1449	22	AAE02535
29	518.5	6.7	1735	22	AAE02535
30	515.5	6.7	837	16	AAK80320
31	515.5	6.7	837	18	AAK26353
32	515.5	6.7	837	19	AAK57659
33	515.5	6.7	837	21	AAV78290
34	513.5	6.7	837	19	AAK53530
35	512	6.7	1487	23	ABG61861
36	510	6.6	1466	22	AAE02534
37	510	6.6	1466	22	AAE02534
38	509	6.6	140	22	AAO05855
39	509	6.6	1464	22	AAK82454
40	508.5	6.6	1418	15	AAK59751
41	508.5	6.6	1418	22	ABR90764
42	508.5	6.6	1418	22	ABR35624
43	507	6.6	2063	22	ABG17147
44	506.5	6.6	1463	22	AAE02532
45	506.5	6.6	1745	22	AAK77793

ALIGNMENTS

RESULT 1	
ABR11808	
ID	ABR11808 standard; peptide: 1435 AA.
XX	
AC	ABR11808:
DT	
DE	11-JAN-2002 (first entry)
XX	
XX	Human BCL9 homologue; SFU ID NO:2178.
XX	
XX	Human: cytokine; cell proliferation; cell differentiation; growth factor;
XX	haematopoiesis regulation; tissue growth; immunomodulatory; activin;
XX	inhibin; chemotaxis; chemokinesis; cancer; tumour; haematopoietic disorder;
XX	proliferation; metastasis; cancer; tumour; haematopoietic disorder;
XX	myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
XX	chronic inflammatory condition; proliferative retinopathy;
XX	atherosclerosis; coronary heart disease; arterial ischaemia;
XX	bone disorder; osteoporosis; vascular growth disorder;
XX	tissue regeneration; wound healing; infection; immune disorder;
XX	cell culture; drug screening; gene therapy; antiinflammatory;
XX	antislammatic; antiarthritis; haemostatic; antiarteriosclerotic;
XX	cytostatic; osteopathic; vasotropic; cardiant; vitruoid; antibacterial;
XX	antifungal; vulnerrary; antulcer.
XX	
OS	Homo sapiens.
XX	
XX	W0200157188-A2.
XX	
XX	09-AUG-2001.
XX	
XX	05-FEB-2001; 2001WO-0803800.
XX	
XX	03-FEB-2000; 2000US-0496914.

27-APR-2000; 2000US-0560875.
 (HYPE-) HYPEQ IM:
 Tang YT, Liu C, Dimaide RT:
 WPL: 2001-457740/49.
 N-PSDB: ABA09052.
 Human proteins and DNA encoding sequences useful for preventing, treating or ameliorating a medical condition in a mammalian subject e.g. arthritis and cancer -

Claim 20: Page 256-257; 196tpp: English.

Sequences ABB10941-ABB12430 represent 1350 novel human polypeptides, and sequences ABA08225-ABA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of identifying compounds which are polypeptides in a sample, and methods of identifying compounds which are polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoietic regulatory activity; tissue growth activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g. by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g. asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g. osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g. of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention.

Query March 97.0% Score 7457.5; DB 22: Length 1435;
 Best local similarity 99.48; Pident. No. 0;
 Matches 1365; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

1 MHSNPKVRSSPGNTGSSPKSKGEVAVRPPTVMSPGNFDLSDKSKGQSSASQSO 60
 41 MHSNPKVRSSPGNTGSSPKSKGEVAVRPPTVMSPGNFDLSDKSKGQSSASQSO 100
 61 PPSGDSKSGHTTKALTPGSSKGLKMGACGAKGKREKSIASDSFVADPSTPHMS 120
 101 PPSGDSKSGHTTKALTPGSSKGLKMGACGAKGKREKSIASDSFVADPSTPHMS 160
 121 DIKSNKADHKSGDSOHTPHSMTPSNATAPRSSTPSHGOTATPEPTPAKTKVYVF 180
 161 DIKSNKADHKSGDSOHTPHSMTPSNATAPRSSTPSHGOTATPEPTPAKTKVYVF 220
 STEMANRAAEVLEGVETIVSFHTONISNNKTERSTAPLNTQISALRDPKPLFQQFA 240

221 STEMANRAAEVLEGVETIVSFHTONISNNKTERSTAPLNTQISALRDPKPLFQQFA 280
 241 PANQDSSQNTKLOPTPTPAAPKPAAPRPLDRSPGVENKLIHSVGSFASSTPLIF 300
 281 PANQDSSQNTKLOPTPTPAAPKPAAPRPLDRSPGVENKLIHSVGSFASSTPLIF 340
 301 DGTGNSTPNNRAVTPVSGSSSSADPKAPPVPSGGEPTLGENPDGJSOLEHRE 360
 341 DGTGNSTPNNRAVTPVSGSSSSADPKAPPVPSGGEPTLGENPDGJSOLEHRE 400
 361 RSLQTLRIQMLFDEKEFTGAUSGGQDMPGLDQKKRPEGLTAMMAQSSLSKGP 420
 401 RSLQTLRIQMLFDEKEFTGAUSGGQDMPGLDQKKRPEGLTAMMAQSSLSKGP 460
 421 GPRDVGAPFGQGRHVDVPSFDEKVPVPSMNSUSTIGPDHLMHTPQJAMKLQOEY 480
 461 GPRDVGAPFGQGRHVDVPSFDEKVPVPSMNSUSTIGPDHLMHTPQJAMKLQOEY 520
 481 EKRKKKQGVVVOQSLQDMVHGHGPRGVYRGPPEYQMTPESEMAHGSTEPSGIM 540
 521 EKRKKKQGVVVOQSLQDMVHGHGPRGVYRGPPEYQMTPESEMAHGSTEPSGIM 580
 541 PHSLEPRGKAPRPNMPSQMLPGFACMINSMEGPNVNFASRGLSGVSMVLGRIF 600
 581 PHSLEPRGKAPRPNMPSQMLPGFACMINSMEGPNVNFASRGLSGVSMVLGRIF 640
 601 DGRNPPGQGTSGSRGRHFPNPGLSBEMVQOOLAKKULGLPRGMBEIRSMENH 660
 641 DGRNPPGQGTSGSRGRHFPNPGLSBEMVQOOLAKKULGLPRGMBEIRSMENH 700
 661 MIPGSRHEPEGNPFRIPPEGLPSRGLPEKIPQMGJGHELEFEEVWPSMKQIV 720
 701 MIPGSRHEPEGNPFRIPPEGLPSRGLPEKIPQMGJGHELEFEEVWPSMKQIV 760
 721 NUNVNGNSQMLPKMKRPAQGEPEMLKIRFGSSDMLPAQKMYPLRGEHPOVETMG 780
 761 NUNVNGNSQMLPKMKRPAQGEPEMLKIRFGSSDMLPAQKMYPLRGEHPOVETMG 820
 781 PPEPLPMSSGPGSSNGLRMLRPPIGPDQRTNSRLSHMPLPLNPSNLTINTAPVGR 840
 821 PPEPLPMSSGPGSSNGLRMLRPPIGPDQRTNSRLSHMPLPLNPSNLTINTAPVGR 880
 841 LGRKPLDTSVAGSVHSPINFLKSPTHOVQSPMLGSPGLKSTQVPSLALMAGA 900
 881 LGRKPLDTSVAGSVHSPINFLKSPTHOVQSPMLGSPGLKSTQVPSLALMAGA 940
 901 AASTKSPVLSAAAPVHLKSPSLPAPSPGTSPPKIFQSPQIPNNKAVLTMASVA 960
 941 AASTKSPVLSAAAPVHLKSPSLPAPSPGTSPPKIFQSPQIPNNKAVLTMASVA 1000
 961 MGNVNSGGPPPTASQASVNI PGLSPSTTYTTPPRTLSQNTLSIMSKMSFAMIS 1020
 1001 MGNVNSGGPPPTASQASVNI PGLSPSTTYTTPPRTLSQNTLSIMSKMSFAMIS 1060
 1021 STEPLV-HDAIKTVASSDQSPARSFNLPSSNNMOMGINTONKRLSINPVVMTLST 1079
 1061 SNGYHMDAKTVASDDQSPARSFNLPSSNNMOMGINTONKRLSINPVVMTLST 1120
 1080 MGMTPLSHSNOMPSINAVGPNIPHGVHMGIGLMSHNPIMGHGSGEPTMVQSGMGTHQ 1139
 1121 MGMTPLSHSNOMPSINAVGPNIPHGVHMGIGLMSHNPIMGHGSGEPTMVQSGMGTHQ 1180
 1140 GPPVPOSPVQVPPHNGSSGUSGSPGSMATPGGAPLTKESNNTVSSADALCKPGGIG 1199
 1181 GPPVPOSPVQVPPHNGSSGUSGSPGSMATPGGAPLTKESNNTVSSADALCKPGGIG 1240
 1200 GPDSTVLNNSMPSVPTDLDLGVIPKATGPIPEFLMSKLISEKFSGLTQVFGKEVIG 1259
 1241 GPDSTVLNNSMPSVPTDLDLGVIPKATGPIPEFLMSKLISEKFSGLTQVFGKEVIG 1300
 1260 KQPGGPGGPGSHMGGMGEQAFRMGLALPGMGGPGVSTPPIFSTAFSMGSHNMPRP 1319
 1301 KQPGGPGGPGSHMGGMGEQAFRMGLALPGMGGPGVSTPPIFSTAFSMGSHNMPRP 1360

QY 1320 AFLQGGMGPHHRMSPAGSTMGQPTLMSNPAAVGMIKCRKSPAGLYTHPVGSPG 1379
 DB 1361 AFLQGGMGPHHRMSPAGSTMGQPTLMSNPAAVGMIKCRKSPAGLYTHPVGSPG 1420
 QY 1380 MMSMGGMGPGPOO 1392
 DB 1421 MMSMGGMGPGPNR 1433

RESULT 2

AAU78460
 ID AAU78460 standard: Protein: 1494 AA.

AC AAU78460;

DT 02-JUL-2002 (first entry)

DE Mouse beta-catenin nuclear localised protein.

XX Mouse; beta-catenin nuclear localised protein; cancer;

XX gene therapy; EST; expressed sequence tag.

XX Mus musculus.

XX WO200224738-A1.

XX 28-MAR-2002.

XX 19-SEP-2001: 2001WO-JP08140.

XX 22-SEP-2000: 2000JP-0287876.

XX (KYO) KYOWA HAKKO KOGYO KK.

PI Akiyama T, Adachi S;

DR WPI: 2002-330014/36.

DR N-PSDB: ABK47631.

XX New beta-catenin nuclear localised protein for diagnosis and treatment

XX of diseases associated with nuclear localisation of beta-catenin e.g.

XX cancer.

XX Claim 1: Page 81-88; 113pp; Japanese.

XX The invention relates to a beta-catenin nuclear localised protein

XX and DNA encoding the protein. The protein and encoding DNA are

XX applicable in diagnosis and treatment of diseases associated with

XX nuclear localisation of beta-catenin e.g. cancer, including gene

XX therapy. The present sequence represents the amino acid sequence of

XX mouse beta-catenin nuclear localised protein.

XX Sequence 1494 AA;

XX Query Match 26.3%; Score 2026.5; DB 23; Length 1494;

XX Best local similarity 36.8%; Pred. No. 2.8e-1107;

XX Matches 602; Conservative 178; Mismatches 491; Indels 363; Gaps 84;

QY 129 DHKSDSDSHSTKTP-----SNATAPRSSTPSHGTTATP-----T 167

DB 183 D-----PAMVTPGLGQTAQLPLSESSAV-----GVHGVQVQVQVQVQVQV 229
 QY 168 PAKGTAKVYVESTEMNKAELKGOVETIVSFHUNISNNKTERSTANILNLSAL 227
 DB 230 PG-KPPSGVYVFTHLANTAAEAVLQGRALSTIAIHQVNPRAKLQA----- 277
 QY 228 KNDPKPLVQUPAPANQDQSSONTRIQPTPIAPAPKAPAPRLLRSP-----GVNKL 285
 DB 278 ---PKVPTPEPLPLN---TESAGTHQSPPLPLPPPIAPQSAIPALPRPPEEDISGL 341
 QY 286 IP-SVGSASSSTPLPDGDTGEPSTNNKAVTPVLSGSSSSADQKA---PIIPVSSSEI 342
 DB 332 AENSVG---AASGGGTGTHPTPTAATANNPILPGQIPGASALILCEATPTGNGGN 389
 QY 343 TISENDOLSOBELHREKSLQTLRDIGMLP---FDEKEPT-----GAGSGP --- 488
 DB 390 LVGS---EGLSKQELHREKSLQTLRDIGMLP---FDEKEPT-----GAGSGP --- 488
 QY 389 -QGNQVLDGPGKKEGEGPQAMMAGSUSLGGPGKIVGAPVQVQVQVQVQVQV 447
 DB 448 ADPPSAPPGGLKTEEPLOSMISOTSLG---GPLEHEVEGTHVQSD ---K 495
 QY 448 PSKNSGSGTIGPDHLDHMTPEQIAWLKQCFEYEEKRKQOVVVQ---QGLQMMVHQH 496
 DB 496 QGNMNMQRLGDSL---TPGVAARKQGEYERKKRKEKQIOLHQRILQIMV---G 548
 QY 507 PKGVV-KGPPPYQMTPESGMAPG-GTE---PESDGINMHSLEPKMAHPNMPN 561
 DB 549 MGGMGRGPPPYHSHKPDQOCAPGAGVALKSPMD---VQDFMOLKPPPPPPPPPP 606
 QY 561 KLPGFAGMINSMEGPNVNPASBPGLSGVSMPLQVRIPIGRNRP---PQGLFS 654
 DB 607 RVPGFGQMSMMEVP---MNAQQRVYHGMAMNELPPIIGSPNPAANAPYRG---G 659
 QY 615 PVRGKRPVQGLSEDFQOQIAEKQ-VGL---PYQAMGDIHFSMEKNKIPNSQRHNP 671
 DB 660 QGEAEFMTPI-VRELLRHQLEKRSKMQRPLQMGSGMSMEKNKIPNSQRHNP 717
 QY 672 GNNPIFP-----RIPV-----EGPLSRKRPF-KQIHQMGSGHLEFQVMP 719
 DB 718 A---MPQGMTGQDGLAGIPMGIEFGGRGLSTPMDQSLKREVDHFKQGN----- 746
 QY 714 SKKGDVNLVNGMSNSQMP-----QKMR-----FAGQPEMLKLR-PQDSME 748
 DB 767 LNNMNVNMMNNMLNVQMTPOQMLMSQKRGQDMQPGQISPEEMAVVANSNGM 826
 QY 759 PAQOKV-VLPDGEHPQOGEYGMGPRPLPLMSQSGSNSJLNLREPTGHRT--- 819
 DB 827 GQFOKMLPSQFPNQGQOQFSQGGQPVQAMQDMQ-----NTIMFSIQSSVPMQIV 860
 QY 811 NSRLSHMPLPLNPSSNPT-SLNTAPPVQGRGLQKRPILISVASQVSHSINPLKSH 869
 DB 881 TARLSHM---PLPASNPFGSVHLAS---NRGLGRPPLDLSISINQMSNPCHLKSTIS 945
 QY 870 QVQSPMLGSPSNLKSIVQTSQALAGPAA---AASIKSPVIGSA-----AASVH 941
 DB 936 QVHSPVLTSPSNLKSIVQTSQALAGPAA---AASIKSPVIGSA-----AASVH 941
 QY 922 KPSLPAPSPGWTSSPKPLQSPGIPFNHKAFLTMASAMGNGVSGSPF---IAS 977
 DB 993 KPSMAVSPGVAVPKRAMPRQVQSKQKQPLISINNSLTGANTQVQALPNAIRNSVA 1062
 QY 978 PASVNTGSLPSTVYTMPEPLISQNLSTIMSKHNSKFAPISSVLYHAIKTVANS 1067
 DB 1053 PPAHSSGIMNPSLEPSTSPPTPSQNLSTIMSKHNSKFAPISSVLYHAIKTVANS 1112
 QY 1038 DSHPAASPPLPMNNMPMGINTQNPRIQGNP---VVPPISTSPHST---QVLSH 1089
 DB 1113 ELLPDR-PLLPFPFPGQSG-----PGISNQDPQMHNNIAAASFPQGMNIPQVLSH 1166
 QY 1090 ---NOMPSPNAVGNIPPHGVPMGGLMSHNPIM---CHGSDPEVPTVQGMFTQGP 1142
 DB 1167 PPTMLPSTFTPLSNIPLRPNAGTGGSSSNMAMVAGGPDLSLNAIPGVFSSQMSHSP 1226

[illegible][illegible]

[illegible]

XX	27	N V	2001	
XX	28	N V	2001	
XX	29	N V	2001	
XX	30	N V	2001	
XX	31	N V	2001	
XX	32	N V	2001	
XX	33	N V	2001	
XX	34	N V	2001	
XX	35	N V	2001	
XX	36	N V	2001	
XX	37	N V	2001	
XX	38	N V	2001	
XX	39	N V	2001	
XX	40	N V	2001	
XX	41	N V	2001	
XX	42	N V	2001	
XX	43	N V	2001	
XX	44	N V	2001	
XX	45	N V	2001	
XX	46	N V	2001	
XX	47	N V	2001	
XX	48	N V	2001	
XX	49	N V	2001	
XX	50	N V	2001	
XX	51	N V	2001	
XX	52	N V	2001	
XX	53	N V	2001	
XX	54	N V	2001	
XX	55	N V	2001	
XX	56	N V	2001	
XX	57	N V	2001	
XX	58	N V	2001	
XX	59	N V	2001	
XX	60	N V	2001	
XX	61	N V	2001	
XX	62	N V	2001	
XX	63	N V	2001	
XX	64	N V	2001	
XX	65	N V	2001	
XX	66	N V	2001	
XX	67	N V	2001	
XX	68	N V	2001	
XX	69	N V	2001	
XX	70	N V	2001	
XX	71	N V	2001	
XX	72	N V	2001	
XX	73	N V	2001	
XX	74	N V	2001	
XX	75	N V	2001	
XX	76	N V	2001	
XX	77	N V	2001	
XX	78	N V	2001	
XX	79	N V	2001	
XX	80	N V	2001	
XX	81	N V	2001	
XX	82	N V	2001	
XX	83	N V	2001	
XX	84	N V	2001	
XX	85	N V	2001	
XX	86	N V	2001	
XX	87	N V	2001	
XX	88	N V	2001	
XX	89	N V	2001	
XX	90	N V	2001	
XX	91	N V	2001	
XX	92	N V	2001	
XX	93	N V	2001	
XX	94	N V	2001	
XX	95	N V	2001	
XX	96	N V	2001	
XX	97	N V	2001	
XX	98	N V	2001	
XX	99	N V	2001	
XX	100	N V	2001	

[illegible][illegible]

[illegible][illegible]

Matches 464: Conserved 462 Mismatches 540: Indels 11: Gaps 822
 UY 40 PPVYMSISGMILILSKSPNPKQKASASQSPSPSKRSQDHPKALSTLQSSILKN 87
 DB 26 PPFAADIMG---AHQDAPKIAHGPACIKATGFAHGGSKDGGPGAGAGFAHGGSK 82
 UY 84 GATNAGKAKIKKRRKRSISALSFQDHPGCTINDISIKERNSAHQIKSQDSQHLFESM 149
 DB 83 LGGPGAGAPAGPAGS-----KDEPG-----EP 135
 UY 147 NAIAPRSSTISHQIAIEETPAUKTFAKYVYVSTPMANKAAFAVKKAVETVSHILQ 236
 DB 106 GATPAPAPAGPAGS-----PAG----- 122
 UY 207 NLSNKLTKRSIAPLNTQISALFNPKPLPGCPANANQNNNNNNHPLGPTGFAFAK 266
 DB 123 -----GQAPGAPAGPAGSRRHGPPAGPAP-----GPGSGAGGPT-----GAGG 163
 UY 267 PAATPKPLDRESPEVKKLIPSWSPAS-STPLPGDTGHNSTFNNKAVTVSSASNSSS 429
 DB 164 PAGHPG---SKDHPGAGAPAPAGPAGSRRHGPGAGPAGP-----SKSDHP 210
 UY 426 ADKAPPP---PVSSDPEPTLGNPKLSQGLFHKRSIQLIKDLYRKFETKFLGA 384
 DB 211 GPPAPAPAPAGPAGSRRHP----- 228
 UY 384 QSSDHPQNVNVLDPQKRPKPLQVAMMAQSSQSLKQDHPRTV-----APGDPGHPV 448
 DB 229 ---GPGAGP-----PAGP-----PSSKDDHPGAGPAGPAGSRRHP 293
 UY 439 PPSHLEKVPISMSNSSTIGPHILDMITFQDAMKILQDFEKKKKKQVYVQV 498
 DB 264 PGPAGAGAPADKQAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 299
 UY 499 DMVYDHPPEVAVGP---PPVYGMISGMAVQ---GTFPSQIMNHSE 549
 DB 297 -----HGTAKKQAGPAGPAGSRRHGPGAGPAGPAGPAGPAGPAGPAGPAGPAG 500
 UY 550 APHPNMISNMRLDPAAMINSEH---GPNVNPASRPILSYVSWGQVKKLITGPNP 607
 DB 451 KQDHPGPA---PPAGPAGSRRHGPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 402
 UY 608 GQGTFSAPGGERFPNPGHLSHPHQQLAEKQLGPPGMAMGKIPESMNMHETPSSR 667
 DB 403 GSRGDMPPGA---PGPAG-----PSSRRGPGAGPAGPAGPAGPAGPAGPAGPAG 442
 UY 668 HMEPCNPPLEPRLPVEGPISP---SRGD-ETKGLTPQMGVRELEPGMVSRKANLV 724
 DB 443 DPGGPAP-----GPAAGPAGSRRHGPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 479
 UY 725 NMGSNSQMTPOKKRFAFAGPEELKLRPQ---GSDMLPAQKKVVLGPHLEAGPEYGM 761
 DB 480 -----PGPPGAPGAP-----PSSRRHGPG-----AGFAHAGPAGSRRHGPG 518
 UY 782 KPTLPMSQPGNSNGLRNLREPTGPKRTSRILSHMPPLINSSNITSLNATVWGL 841
 DB 519 ---PGAHGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 568
 UY 842 GKRTLLSVAGNVSPGPNLPSPTMHWVOSTMLASNSGNKSYVTSGLAMATIAA 901
 DB 569 -----HGGPAGP-----AGPAGSRRHGPG-----AGFAHAGPAGPAGPAG 601
 UY 902 AASTKSPVLAASAATVHLKSLT-----PAISDQWISSPPPLQSSSTLTKRAV 954
 DB 602 SRGDHPGAPAPAGPAGSRRHGPGAGPAGPAGPAGSRRHGPGAGPAGPAGPAGPAGPAG 664
 UY 955 IMASIAMGNVESGAPPTLASQASVNPPOSTPSTPYIMPTETLSQNTSLMSRR 1013
 DB 662 PNVAGPAGPAGSRRHGPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 703
 UY 1014 SKFAMSPSTFVHDAKTVASSSDSPYAKSNLNSMNMPPGMLNCPNPS 1073
 DB 704 ---PAGPAG-----GSRGDPG---GAGPAGPAGPAGSRRHGPG---GAGFAHAGP 743

UY 1074 MPTLSMGMLOETSHSNQMSNAVGNLP---PVIQMLMSHPTMHSQHP 1127
 DB 744 -----PSSRRHGPG---GAPGAPAGPAGSRRHGPG---AGFAHAGPAGPAGPAG 789
 UY 1128 ---PVIQRMQ---PQGPFWQSTQVQPPHNGSNQNSF---PQMGTEGELQK 1179
 DB 790 GAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 847
 UY 1180 PSLHVSNAIAALCKKQAGPAGPSTVLANSMSVPTDQDQVIRGAGLIEFHSK 1239
 DB 848 PAGPAGSRRHP---PGPAGPAG-----GAPG-----PSSRRHGPG 877
 UY 1240 HSEKPSQLIYPRGGEVGRKQVQNG-PGSHMQAMHAGAP---KMLALP---GMSG 1294
 DB 878 -PVAAGP-----PAPGAGSRRHGPGAGPAGPAGSRRHGPGAGPAGPAGPAGPAGPAG 930
 UY 1295 PGVGTDLHPLT-----APSMGHNMPFATLQDMMGPHHBMSPAGSLMGGPT 1346
 DB 931 GPGAGG-PAGPAGSRRHGPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 972
 UY 1347 LMSNPAAGMLDCK---GGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 1404
 DB 973 ---GAPG-----PSSRRHGPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 1016
 UY 1405 MAADVGMGPGSGGNGP 1421
 DB 1016 SRGDHPGPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 1031
 RESULT 10
 AAR47241
 ID AAR47741 standard: Protein: 1065 AA.
 AC AAR47741:
 XX
 D7 07-SEP-1994 (first entry)
 XX
 DE Collagen like polymer latex encoded by clone p19 a176.
 XX
 KW Recombinant: collagen-like polymer; (192 amino acids; molecular
 KW fiber film; coat; coat sequence; collagen; mammalian; monodimer;
 KW hydrogel; intercalin; linker; collagen suspension; latex and body.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Protein 34..285
 FT /Label: p19p3
 FT Protein 286..507
 FT /Label: p19p4
 FT Protein 508..789
 FT /Label: p19p3
 FT Protein 790..1041
 FT /Label: p19p4
 XX
 PN M0910154 A.
 XX
 PD 27 MAY-1993.
 XX
 PP 04 NOV-1992; 92NW-03809485.
 XX
 PR 12 NOV-1991; 91US-0791960.
 XX
 PA (PROD) PROTEIN POLYMER TECHNOLOGIES INC.
 XX
 PI Cappello J, Ferrara EA
 XX
 DR WPI: 1993-182496/22.
 XX
 PT High mol. wt. collagen like protein polymers capable of being
 PT produced in unicellular microorganisms

[illegible][illegible]

RA Wang Z.-Y., Massarini D.A., Welstock G.M., Weissbach L.
 RA Williams S.M., Woodson T., Worley K.C., Wu D., Yang S., Yao V.A.,
 RA Ye J., Yeh R.-E., Zaveri J.S., Zhao M., Zhang G., Zhao Y., Zhou L.,
 RA Zhou X.H., Zhou F.N., Zhou W., Zhou X., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.
 RT "The genome sequence of *Drosophila melanogaster*."
 Science 287:2185-2195 (2000).
 IR EMBL: AF004575; AAF50995.2;
 IR Flybase: FB00041617; CG19635.
 IR InterPro: IPR002052; N6_Mtase.
 IR InterPro: IPR002965; P-rich_extensn.
 IR PRINTS: PR01217; PRICHXENSN.
 IR PROSITE: PS00092; N6_MTASE; GNKNWN.1
 SQ SOURCE: 1047 AA; 10615 MW; 76370:AAAZ0736A2 PRO44.

Query Match 7.38; score 559.5; DB 5; length 147;
 Best Local Similarity 24.78; prot. No. 61e-24;
 Matches 328; Conservative 116; Mismatches 448; Indels 452; gaps 84;

UY 198 ETVSPHILNINNN-KTERSTAPNTQISALRNDKRLPQUPAP-----ANQANSSNIR 254
 DB 12 ESYTFHFHYVYVPPNPPPSAPAP-----PPKPPPPPPPPPTTTPPT 61
 OY 254 LGPTTTPAPAPKPAAPRPRLDRESVEENKL-----TSVGSVA-----SSTPTPD 601
 DB 62 TTPPTTTPPPPPSAPPP-TPATPVWYPLPYGNAPQPPPPSHLSNSKRRPP 119
 OY 402 GTGNSTPNNKAVTPVS-----QGSNSSADPKAPPPPPSSSGPPTGGNPKLSNQL 356
 DB 120 -GSGMPNNSAPVTPATGTPYVPPNQQNPPSSQTPNSSETPVPAWMPVPPNQT 177
 OY 357 EHKRSTLQTLNDLQRLPLPKKEFTGAQSGQPPQNGVLDPPQ--KKRPRTGAMMAQSG 414
 DB 178 GH-----PPSSQ-----APTNSGVEPPALWNPDPAP-----PN 206
 OY 415 SLCKGPRGTDVAPFPQGGHROVPSPDEKVPSSMSQSTTPRLLAMTPVQIMLK 474
 DB 207 PGGNPPSSQTPNNGFP-----PPATWSPS--VPTSHDPPPSGA- 247
 OY 475 LQGFVEEKKRKHQVAVVQCSLDMMVHOGPPGVKCP-----PPVMTSEKMAP 529
 DB 248 -----PPNSGVPPLATCTPPNQQNTGQVQ 276
 OY 530 GTEFPNGINNNP--LPPKQ--MAFHPPNPPSSQMTPEFAGINSEMGIVNPPVASK 584
 DB 276 SHKPPNNSAPVTPATGTPYVPPNQQNPPSSQTPNNSL-----FQSTHPPR 324
 OY 585 PGLSTGVPMPVPPKPLGKPPPPGQGFSGPPKRRFPNQQLSSEMPDQGLAKGLAP 644
 DB 425 P-----PPGLQD-PPGSGMPNLSQIYP-PSMG-----WTP 352
 OY 645 P-GMAEGEIKESMMENMIPGSGHMFENNNPFP-KHVEGLSPS-KQGFYKQ--LPP 659
 DB 453 PGGSLTQGGP-----PQS--LPPNSVLAQSTP--PLKPSNGGHTPPSLPP 398
 OY 700 QMGGRLEF-----GVPSSMGKGDVNIIVNMGSSNMTPQ 735
 DB 399 NPAFLFDHQLKLDHQQHQLLMWYTPHPVAGFPDS-----ELLGPHLGGSLTP 453
 OY 736 KMFPAVAGPHEMKLR--GSMMLPADQKRYPLDGGHQQVYMGPRPTPMQAT- 700
 DB 454 ---NMLPPGSTPLPPKPPNGGHP--PGSK--PPVGLYPPSTLWMD-PLGLDGGH 504
 OY 791 ---NGLNLEKPTG--PQRTNSRSHMP-PLPLNNSNPSTNAT--PVQ 849
 DB 505 PPGSLTPPNTGL-----PPSTPTPNSPNDGHPNQGKPPNGLYVSTGWTTPSPPT 559
 OY 839 RGLGKRPDLSVAVSVYSHNINFLKSPIMHVGSPMLASPSNLKSPQFSLANMLAG 898
 DB 560 QG-GHPGSLPPNTGLPPGSLTPPLSPN-----QIGHPPGSQKPPN-----VWY 604
 OY 899 PAAASAKSPVAVGSAASVIVHKSLTAPSGWTSSKPPQSTPTTPNNKAP----- 954

DB 605 PPSTGWLPPSGPLTQSGHPGSLTPNTGLD-PPSLPLPPNPGSLTPNTGLSPSTP 664
 OY 955 TMASPMALGVNENGGPPPPASPVANLQSLPSTPYIMPEPTISQPLSSMSBS 1014
 DB 664 PLSGPNGLGHPGSGKPPNSSTYPPST--GSTPSTLLGKAP--KSL----- 709
 OY 1015 KFAFSTPLVYHAKTVASSIDSPSPSPNPSMNMIMQINONPISAPNVIR 1074
 DB 710 ---LPSNTGLPSSST-----PPASPN-QGCHPPG--SKPTNSG---LYP 747
 OY 1075 PT-LISMATQTLSHSNQPSIRAVPN--TPHRY-PMG-GLKSHNPIMHNSQPP 1128
 DB 748 STWMLPPSG--PLTQGGHPG-SLPPNTGLPSTSPASGSLQIGHP--GSKRP 709
 OY 1129 MVPPGRRGFTQGGPPVVS--PVGVPVPPNPGSGQ-----GSPQPMPE-PPRHLGAP 1180
 DB 800 NV-----GI--YPTSGTTPPLPPNMGGHPGSGPPVNVGLYPSSTWMPSSG 851
 OY 1181 SNLPPSSAALCKRPPGGGPPSTVAGNSMPSEFTIDPLEVYVNAATLPPPLSK 1240
 DB 852 GHP-----GSGVPS-----NSVLPNS----- 870
 OY 1241 PSKPSQTLQYTRGV-----PKKQPGGPGPSHMGCMGVALKMGALGMA 1294
 DB 871 PPLPPNQGHPHPSVPPNTGLPPGSLPPGSDNGLGHP--SKPPNSCTYPSTGS 928
 OY 1294 -----PGVGTID-PLGLTSMGHNIMPRALQGMTPHRRMSIA 1358
 DB 929 LPSGLPLQGGHPGSLTPNTGLPSSSTP-----PLKPP--NQRHPPSGV--TPN 977
 OY 1359 SIMP-----GPTLMSNAAVGMIPKURPACLYTHGVPSSIMGSMGCMG 1391
 DB 978 SVLPPGSLPPGSLQIGRVGS-----QKPAASKIPPTTSPQNTNV-----IK 1024
 OY 1392 QNIMPP 1398
 DB 1025 QPVKNPP 1031

RESULT 7

UYVE67 PRELIMINARY: PRT: 2703 AA.
 AC QVVE67;
 DT 01-MAY-2000 (FEMBLrel. 13, Created)
 DT 01-MAY-2000 (FEMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (FEMBLrel. 21, Last annotation update)
 DE GSA protein.
 GN GSA OR G7467.
 OS *Drosophila melanogaster* (Fruit Fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; *Drosophilidae*; *Drosophila*.
 OX NCBI_TaxID=7227;
 RP SEQUENCE FROM N.A.
 RC STRAIN BERKELEY.
 RX MEDLINE: 20196006; PubMed: 10741142;
 RA Adams M.D., Gelinkler S.E., Holt R.A., Evans C.A., Corayne J.D.,
 RA Amanatides P.G., Scherter S.E., Li F.W., Hoskins R.A., Galle R.F.,
 RA Guoqiao R.A., Lewis S.E., Richards S., Ashburner M., Brinkston S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.H.C., Blazer V.G., Champ M., Pfeiffer R.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Holt C., Nelson C.R., McKus G.L.D.,
 RA Abrait J.F., Adleyah A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Haller R.M., Hsu A., Bakendale J., Bayraktaroglu C., Beasley E.M.,
 RA Heesoon K.Y., Henson P.V., Beyman B.P., Blandart J., Brashkov S.,
 RA Burkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Centur A., Chahud L.,
 RA Cherry J.M., Cawley S., Dahlke C., Davernport L.H., Davies P.,
 RA de Fabis B., Decher A., Deng Z., Mays A.D., Iwak D., Dietz S.M.,
 RA Dodson K., Dong L.F., Downes M., Duan-Rochia S., Jambry H.C., Jann P.,
 RA Durbin K.J., Evangelista C.C., Foray C., Fortera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gang N.S., Gellart W.M., Glasser K.,

[illegible][illegible]

[illegible][illegible]

[illegible][illegible]

```

Db 1412 GGP-UPPIOTVAGG-----PPAGTAPGAPSSAVPTGPSSAGTYYG- 1472
QY 1256 EVPGKQKQY-----GPGPESHMGMM-----PAGPAPMIALP 1289
Db 1414 ---PPIGSGPKRHDPFKQSPYPTGYMARQYTGAWSGTQYKQYTSIAFWMTAP 1410
QY 1290 GMGGTVGTPIDPLGAVS-----MGHNDMPAPVAGQMMGTHHMMSPAGS 1399
Db 1541 PGGAAPPGAPGPGPPLQGFACVAGVMDQIRYPPQGPPEPPPGQGG- 1586
QY 1440 TM- PPGPTLMSNPAAAVOMIPKIDRPPADLYIHFG-----PVGSPHMMSSAPGMMTHQV 1392
Db 1587 QVAGPFGQPPVQAPFGVMA-QNNWQ-FAVSGT APFSPHMPSPSTFGQNNHMGMPVAGQ 1444
QY 1393 NI- ---MLPQMRPGMA-----ADVENMGF-----NPGN 1424
Db 1644 QSGQGGVPPVPPQVQVSHGVPSHPLQVQVGHQVKKPYAMPPTPSQVAGV 1704
QY 1424 MM 1425
Db 1704 MM 1705

```

RESULT 12

```

Q60467 PRELIMINARY: PRI: 1840 AA.
Q60467

```

```

Db 01-NOV-1996 (TREMBL:01, Created)
Db 01-NOV-1996 (TREMBL:01, Last sequence update)
Db 01-JUN-2002 (TREMBL:21, Last annotation update)
Db Pro-alpha-1 type V collagen.
Db Cricetus leonticaudatus (Long-tailed hamster) (Chinese hamster);
Db Eukaryota: Metazoa: Chordata: Granata: Vertebrata: Euteleostomi:
Db Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Cricetidae:
Db Cricetus.
Db NCBI_TaxID:10040;
Db 11
Db SEQUENCE FROM N.A.
Db MEDLINE:92105142; PubMed 172213;
Db RA Greenspan D.S., Chong W., Hoffman G.G.;
Db *The pro-alpha-1(V) collagen chain: complete primary structure,
Db distribution of expression, and comparison with the pro-alpha-1(XI)
Db collagen chain.*
Db J. Biol. Chem. 266:24727-24733(1991).
Db EMBL: M76730; AAA37002.1; -
Db InterPro: IPR000087; Collagen.
Db InterPro: IPR000885; Fib_collagen_C.
Db InterPro: IPR001791; Laminin-G.
Db InterPro: IPR001230; Procollagen.
Db InterPro: IPR001129; TSPN.
Db Pfam: PF01410; COLF1; 1.
Db Pfam: PF01491; Collagen; 18.
Db Pfam: PF02210; TSPN; 1.
Db ProDom: PD000007; Collagen; 2.
Db ProDom: PD002078; Fib_collagen_C; 1.
Db SMART: SM00348; COLF1; 1.
Db SMART: SM00282; LameG; 1.
Db SMART: SM00210; TSPN; 1.
Db EROSITE: PSD0294; PRENITATION; UNKNOWN_1.
Db Collagen.
Db SBOEMME 1840 AA: 184174 MM: 3255821EF6AC75E0E4;

```

```

Query Match 7.1% Score 549.5; DB 11; Length 1840;
Host Local Similarity 22.7%; Pfam No. 5,6e-245;
Matches 454; Conservative 110; Mismatches 499; Indels 591; Gaps 85;

```

```

QY 248 PPAPAMQNNSSQ-----NRIQPTP-----PIPAAPKPAAPR-----FIDR 276
Db 249 PDITQSGPNDVEYVPGDGFYFYFYFYEDPEIDKEIAPYOKPEAAFEIIVHEDG 308
QY 277 ESPGVENKLIPSVGSPASS-----TPLPDGTG-----FN SIP 409

```

```

Db 309 TQPPAPVVEHISPPACKEDDQSGDYVPIDVYTPAVYHAYGVEVNDIDPTN 468
QY 410 NNRAVTPVSGSSSSALPAPPPVSSSGPTLTENPGLSLGOLFERRISGLTKR 469
Db 469 DSGAEVPTSLITTSINSD-APSP-----GDKIDLADEETL -ETIKRL 410
QY 470 QRMLE-----PDEKEEFGAQS-----GRRQUNNG-----VLD 496
Db 411 PENYIDYPPDISNSNSPPETIGPMRANGDTIYVLIIGIKGKQKQKQKQKQKQK 470
QY 497 GPQKRPAPVQAMMAQSLSLQGP-----GRITVQAP--EPPQSHQVPSSEMYVPSM 498
Db 471 GP-PGPRAPAL-----PGRPLTGPFGVQGPGRGPRAPRNLPGAKMLDGHGTL 529
QY 451 -----NSQSITIGPHILHMTPEQVAMIKLQGHETHEKKKQGVVVVQVSG 498
Db 521 MLMLPFGGAGDAGSGKP-----MVSAQ-----ESLQALITQVQARALR 561
QY 499 -----DMVHGHGTRVVRQTPGYOMTPEGMA----- 527
Db 561 GPAPKMLTIGRPGMGRVNSGKIGRQGM--GPGGKRG-VQGPFGTGRKRGKQKQKQ 617
QY 528 ---PQTEPESD-----GIMHNSLP-----PGMA 556
Db 618 GAKMPQQTGCKEDRGEDLAGLPGKGRKQKQKQSPAPLPGDQKRGKQKQKQ 676
QY 551 PHEMPRSQMKL-----PFGAGLT-----NSMRGT 575
Db 677 -PQEPKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 734
QY 576 -----PVPVNSAPKGLSCVSWPMDVPRTPGGRNFGPQGLTSSQSGEDTNGHT 627
Db 735 GPQCATLPGPKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 786
QY 628 SEDMPQQLAEKQGLPFGPM--ENLR-----PSMEKMYLNSQKHM 626
Db 787 -----PLGYPQVQKAKQGLKGLKQKQKQKQKQKQKQKQKQKQKQKQ 844
QY 671 PQNNPFRKIPVRQDIPSPSGDTFKGLPYQMG--GRELEGGVY--SKKQVNV 721
Db 845 PFG-----PR-GEDPSPKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 888
QY 722 LANNMGSSMTQYKMEFAGAGFEEMKILRQSSDMIPAGCKVNLPGCHVQVQVGM 781
Db 889 PFGPFGAGCE-----KQKGTFG--KRG-----PRQKQKQKQKQKQ 925
QY 782 RPLTPKSGQKSGNSGLNLEPDTGPDQKNSRLSHMLPLINSSNLSIN TAPVQK 840
Db 926 RG-LTGKPGKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 975
QY 841 LGKKPLDLSVAGSVDHS-----PGLNPKSPY-----MHVQSPMLQSPSG 881
Db 976 LPHGPGQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 1035
QY 882 ---NKKSTQTSQALAM-----LAPPAASATK-----STPVLS 912
Db 1036 KESTKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 1095
QY 913 SA-AASPVHLKSTSLPATSQWTSSPKPLQSTGPIENKAPLTMAS PAMLVNVSQ 979
Db 1096 PAGAAGPLGT--PGKQKQ-----QGRPGTAGKQV--PGKQKQKQKQKQKQKQ 1147
QY 971 ---PPPTASQVAVNIVGSLPS--STVYIMPEPTLSNPLSLMSSKSPAMSSITP 1024
Db 1148 AGVPGPGEKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 1191
QY 1024 LYHDAIKTVASSSDSPPAS-----PNSKNNMG--MGNI 1360
Db 1192 -----QHGSGADGEGPQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 1245
QY 1061 QNPRTSGPNTVVPPTLSTMGKQPLSHSNQMSPPAVG-----PNIHGVVHG 1110
Db 1246 DWQMGPGKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 1305

```

[illegible][illegible]


```

DB 1065 KDPASGCGPPGKGEPSGCGRTGCVGCKGPPGPPGK-----PAG 1109
QY 1197 GPGGPGSTVGLGMSMSVFLDPLDQEVIRGAGTGFEPDLSKLLPSKPSQLLAFFHGE 1256
DB 1110 KPGGPGSTVGLGMSMSVFLDPLDQEVIRGAGTGFEPDLSKLLPSKPSQLLAFFHGE 1256
DB 1257 V-----GPRKGPAGP-----GPRPSHMOGMMLEDOALPRMGLALDGMGPGGVGGLDPLLA 1407
DB 1147 AGGPGPGPGGPGVGGPGGLGPGGLNHLGKGGKGLKAGSGLLEMLGPGHMG-----GCKRGE 1202
QY 1408 ---PSMTGHNMPGPAFLGQGMMPHHRMSPAOSFHPGPTLMSNPAVAAMTCK--- 1461
DB 1203 KGPBGSGTISPPGLPG---EKGTPGPGPKPGPPGPGAGATK-----AAKGLHPGPGP 1252
QY 1362 -DPCGA---GLYHPGVGSPGCMKMSMGPGQGNIMIPGK----- 1401
DB 1253 GDRPGPGPKPGPGPGPGPGVNDLKGDPG---LGLPGPGSGRPGPGPGAGPGAG 1411
QY 1402 PRGMAADVCMGSGSGGPGNG 1422
DB 1312 KDGKGGHGLGPGPLPG 1332

RESULT 14
QY 09103 PRELIMINARY: PRT: 1840 AA.
AC 09103:
DT 01-OCT-2000 (ITEMBRO1, 15, created)
DT 01-OCT-2000 (ITEMBRO1, 15, last sequence update)
DT 01-JUN-2002 (ITEMBRO1, 21, last annotation update)
DE Alpha 1 type V collagen.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclerogastri; Muridae; Murinae; Murus.
OX NCBI_TaxID:10116;
RN 111
RP SPOURCE: FROM N.A.
RC STRAIN: SPRAUE DAWLEY;
RX MEDLINE:20428740; PubMed:10852920;
RX Chetaniousov M.A., Kolibulum K., Tyler W.A., Stahl R.O., Carey D.J.;
RX Schwann cells synthesize type V collagen that contains a novel alpha
RX 4 chain, molecular cloning, biochemical characterization, and high
RX RT affinity heparin binding of alpha4(V) collagen.
RL J. Biol. Chem. 275:28208-28215(2000).
DR EMBL: AF272662; AAF76433.1;
DR InterPro: IPR000872; collagen;
DR InterPro: IPR000885; fib-collagen;
DR InterPro: IPR001791; laminin;
DR InterPro: IPR001240; prenyl_site;
DR InterPro: IPR000508; signalase;
DR Pfam: PF01410; COL1C1.1;
DR Pfam: PF01391; collagen;
DR Pfam: PF02210; TSPN.1;
DR ProDom: PD000007; collagen;
DR ProDom: PD002078; fib-collagen;
DR SMART: SM0038; COL1C1.1;
DR SMART: SM00282; LamG.1;
DR SMART: SM00210; TSPN.1;
DR PROSITE: PS00294; PRENYLAT-N; UNKNOWN;
DR PROSITE: PS00761; SPASE_13; UNKNOWN;
KW Collagen.
SQ SEQUENCE 1840 AA: 18986 MW: 403855FF886926 GCKC4:

Query Match 7.0% Score 540.5; FR 11; Length 1840;
Best Local Similarity 22.7% Prod. No. 1.8e-22;
Matches 334; Conservative 122; Mismatches 576; Indels 599; Gaps 85;

QY 109 IQDQPGTPIINDSDIKESNSAHILKSQSLHTPSMTSPNAIPASSIPSSGATLAFITF 148
DB 280 IDPDGPKKPAFASO-KVEAA-----KRTTEVPEDGTGPPSAP--LVPELSQLAKEDNP 443
QY 169 A-----GKTAKVVVYVSTEMAKAAEAVLKQVETIVSEHTONTNNKTKFSAPENLQI 224

```

```

DB 332 GIDGVYVPLDIDYYTISPEEDRYGCVENHPQPT-----NPGSAEHLTSTSV 480
QY 225 SALRNDPKPLPGQ-----PPAPADQDN-----SSQNIHQPTTPEALP 265
DB 381 TSSNSPAPSPGPKRIDLAGFEETFKLDENYVPTPTDSTNSV-----SF 429
QY 266 KPAAPPLDRES-----PQVE--NKLPSVGSASSPPLPPDQDGNSTPNNAVPT 416
DB 430 SEIGPGMFMQDTIYEGTGGPKGKQCKEPALFERGMLFHPGPGHAGLPG---PF 485
QY 417 VSQGSNSASADP--KAPPPP-----VSSGEPPL-----GPNFGLSGBOLHRE 460
DB 486 GTTGPGTCQMGDVEGKPPGPPHPLPGAKLPGTGLMLLFFPRGAGLJAGSGPWASAF 545
QY 361 KSLQTLKDTGRLPF---PDKETFLVQ-----SGTPQDNVGLDVG-----KRPKRP 405
DB 546 SQQAALLQQAFLALKGPAPAPMHTGPRGPMGVSQSGKGPEDGM-GVQGPVAVGPPV 604
QY 406 I-----GAMMAQSQSLAKGASPTD-----VAPGPGGHRIVTFSPDEMVPSMS 452
DB 605 TCKPGRRRAGSDGARGMGDTGPKDGPDPGLAGLP-GKKCHRG--DGGSGATPGGLG 660
QY 453 QSGTIGPHLDHMTPEOLAMLKQDFEYFKKKQDEQVVVVOO-STQMMVHQHGRGVV 511
DB 661 DQEGKDD--GVGPGPLP-----GEGPGHGLG 687
QY 512 ---KRPPTPGMTTSPBQAPAGTETPSSGTLNPPSLP-PRGMATDINHGSMKIPGA 566
DB 688 PKGPGPGPPGVNTGNG-----GPRKGNVPLQDEPDPGQGNVAGQ---LHVG 737
QY 567 GMIINSEEDPNNVNASRGLSSVSPDLPVKIPKSRNPPGGLFSGHGRKGRPPNG 626
DB 738 GAGP-----PKKPLGKGLGLQMGVAKKATVGP--GKRPTRGKSLGQI-----PGVQ 786
QY 627 LSEHMPVQOLAKKQGLPQGM-----KQIR-----PSMNNKMLPSQKRM 669
DB 787 -----PIGYPGPGVKMGMLKGLKGLKGEKGGHPLGPKDMGKIKLKEPL 833
QY 670 PRGNNTTPRIVPGGLSKSLGPRGGLPGQPG GGLERPMV-----SGKRVV 720
DB 834 GPGG---PR-GHDPGSGKRGCGNDGPGGLSGPGEKGLD-VGLDQVPGKQPGK 887
QY 721 NLNVNNGSNQMLPKMRBACAPDEMLKTRGGSQMLPAQCKVHLPGEHVGGEYMG 780
DB 888 GPGGPGAGG-----KQCKTFS-----PRGKGTGGRGK-----G 924
QY 781 PRFTLPMSPGPNSSGLNKLKPLTGPQRTNSRLSNMPLPLNSSNPLSLN-LAPVQV 839
DB 925 PRG-ITGKPGKQNSQGG---IAGPYGREG-----ENYGGTGGPGKAPVMPKRD 974
QY 840 GAKKRLDLSVAGSVHS-----PGINLKST-----MHVQSLMANSNG 861
DB 975 GLGHPGPGGTEFGOKTPGPPGPGVVGPGTGTGCMGEGHGLPGPGGEGALVAA 1044
QY 882 ---NLKSPVPSQLAGM-----LAGVAASAK-----STPLV 911
DB 1035 GKEGKKGDPGPGAGLPGKAPVNLKRPVGRGLPGVGLAGLKGSLGPPGPAASPR 1094
QY 912 GSA-AASPVHLEKSLPAPSPMTSSPKPLQSPGLPPNH-----KAPLTPASP 959
DB 1095 GPVAGAPPGI--PGRPV-----GSHVPGAGEKVPVGEHNGGAGAGLGGVGLASP 1147
QY 960 AMLQNVESGAP--PPVTSQSPAVNPQSTLS--STPTMPPPTLSQNTSTMSRMSK 1015
DB 1148 A-----GPGVPGDEMDKGGTGFPGQKSGKGRGPGGPTLDPGPTD----- 1191
QY 1016 FAMPSTPLYHAKITVASDDSPAS-----PNISSMNMGG 1095
DB 1192 -----GPGSCAGLNLNPRKQOGLFQCKGLGDSKGTGPGVGLGGLASP 1247
QY 1056 ---MGINTUNPRISQNPVVPMTLSPMKMTGPISSHMSQNTSPNAG-----NRI 1192

```



```

Db 1085 GEAGLEGTGCL-----TGPAGQKGEPSNQLPMSAGKTEPDIJMKTPHDF 1180
QY 1111 PMLSHNPIMH-----GSDPPMVP--GGHMGF-----PGSPPVQSPVQVFFHNGS 1159
Db 1131 PQAKGDKRSKGVGPPVGLAGSPGITHSKQEGGFMPPGFQVQPGIAGSPHAT EATPK 1147
QY 1160 GGQGSFPGMGHPG-EGHLPSPNIPQSSADAALTKQGGPGGPPSPFVLTNSMSVFTDP 1216
Db 1188 GDRGP-OGDPGLPQIHPGFMG PGLP--GHDGVKQDKGNPGMPCV--GVVGPRTDP 1238
QY 1219 DLQVTPRGATGPEPDLSTTPSEKPSQTLQYFFKGVPRKQFQGGHGFSSHMLGMMG 1278
Db 1239 GFQGM--PGIGSPG-----ITGSKGDMG---FPG-VPGFQGPQGL-PGLQALKEIQQ 1284
QY 1279 EQAPRMGLALPGMGPGPVPTPDIPLGTAPEPMHNPMPAPVLOQMMTPHHGMSNFAQ 1338
Db 1285 DQGVPGAKGLP--GVPGPGIYDIIKG-EGGLPG--EGCPGL--KILGG----- 1327
QY 1339 STMGGQPTLMSNPAAAVQMI PKQDRGPAQLYTHPGVWSPGM--MMSMGMMGTPQUNIM 1396
Db 1328 --LPG-----PKGGQVTVGLVGTGTPPGCTPGFDGAPQKQGMCP --- 1364
QY 1367 PPGMRPRGMADVGMGSPSGGPNPG 1422
Db 1365 AGPTGPKRFPQGPQGDIPMSMSPPG 1390

```

Search completed: February 28, 2003, 16:08:41
 Job time : 112 secs



Genotore version 5.1.4
Copyright (c) 1993 - 2003 Compugen Ltd.

00M protein - protein search, using SW model

Run on: February 28, 2003, 16:04:04 : Search time: 17 seconds

(without alignment)
3479.132 Million CPU updates/sec

File: US-09-915-543-15

Perfect score: 7692

Sequence: 1 MISSMPKXKSSISNCTQSSP.....ADVMDKPSGKNNIENMDF 1426

Scoring table: BLOSUM62

Gapop: 10.0, Gapext: 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112932

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 108

Listing first 45 summaries

Database: SwissProt_40.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Protein Name
1	605.5	7.9	1685	CA54_HUMAN	EC24000 homo sapien
2	600	7.8	1763	CA24_ASCTO	P27393 ascorbic acid
3	571	7.4	1669	CA14_MOUSE	P02463 mus musculu
4	565.5	7.4	1670	CA54_HUMAN	Q01495 mus musculu
5	555.5	7.2	1838	CA15_HUMAN	P20008 homo sapien
6	552	7.2	1690	CA44_HUMAN	P54426 homo sapien
7	543.5	7.1	1466	CA13_HUMAN	P02461 homo sapien
8	545	7.0	1669	CA14_HUMAN	P02462 homo sapien
9	530	6.9	2944	CA17_HUMAN	P02468 homo sapien
10	528	6.9	1758	CA24_CABEL	P17141 consoliated
11	523	6.8	1678	CA64_HUMAN	Q14631 homo sapien
12	517.5	6.7	1464	CA13_MOUSE	P08121 mus musculu
13	516	6.7	1453	CA11_CHICK	P04457 gallus gall
14	506.5	6.6	1650	CA2B_MOUSE	P04459 mus musculu
15	505	6.6	1464	CA11_HUMAN	P02452 homo sapien
16	504.5	6.6	1464	CA12_HUMAN	P02458 homo sapien
17	503	6.5	1804	CA18_MOUSE	P04445 mus musculu
18	502.5	6.5	1459	CA12_MOUSE	P04441 mus musculu
19	501.5	6.5	1453	CA11_MOUSE	P11987 mus musculu
20	501	6.5	1453	CA11_MOUSE	P11987 mus musculu
21	500.5	6.5	1603	CA1F_HUMAN	P19442 homo sapien
22	498	6.4	1460	CA11_CANFA	Q08517 caris fabli
23	496	6.4	1496	CA25_HUMAN	P08507 homo sapien
24	490	6.4	1049	CA13_BOVIN	P04458 bos taurus
25	488	6.3	1262	CA13_CHICK	P12105 gallus gall
26	488	6.3	1806	CA1B_HUMAN	P12107 homo sapien
27	484.5	6.3	1712	CA24_HUMAN	P08572 homo sapien
28	471.5	6.1	1707	CA24_MOUSE	P08122 mus musculu
29	463	6.0	1516	CA1H_HUMAN	P03466 homo sapien
30	460.5	6.0	1364	CA21_BOVIN	P02465 bos taurus
31	455	5.9	1362	CA21_CHICK	P02467 gallus gall
32	449	5.8	1758	CA14_CABEL	P17141 consoliated
33	449	5.8	1775	CA14_OROME	P17141 consoliated

44	448	5.8	1372	CA21_RAT	P04466 ratios betv
45	444	5.8	2142	CA12_HUMAN	P04464 homo sapien
46	442.5	5.8	1466	CA21_CANFA	Q01492 caris fabli
47	442	5.6	1466	CA21_HUMAN	P08123 homo sapien
48	428	5.5	2167	SHR1_KAT	Q08548 ratios betv
49	425.5	5.5	1855	CA21_RANOA	Q04460 ratios betv
40	425.5	5.5	1902	SMF1_HUMAN	Q14497 homo sapien
41	424	5.5	1372	CA21_MOUSE	Q01449 mus musculu
42	423.5	5.5	1527	CA1B_MOUSE	P04461 mus musculu
43	412	5.4	779	CA11_BOVIN	P02453 bos taurus
44	410	5.3	1456	CA21_CONOV	P02464 consoliated
45	410	5.3	2161	SHR1_HUMAN	Q08566 homo sapien

ALIGNMENTS

RESULT 1
CA54_HUMAN STANABRO: PRT: 1685 AA.
ID CA54_HUMAN
AC P29400: Q16126: Q16006;
DT 01-08-1992 (rel. 24, created)
DT 01-FEB-1994 (rel. 26, last sequence update)
DT 15-JUN-2002 (rel. 41, last annotation update)
DE Collagen alpha 5(IV) chain precursor.
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eulauria; Primates; Catarrhini; Hominoidea; Homin;
OC NCBI_TaxId=9606;
LN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:94165049; PubMed:8120014;
RA Zhou J., Leinonen A., Trygvasson K.;
RT "Structure of the human type IV collagen COL4A5 gene";
RI J. Biol. Chem. 269:6608-6614(1994).
RN [2]
RP SEQUENCE OF 1-510 FROM N.A., AND VARIANT AS CYS 521.
RX TISSUE: Kidney;
RP MEDLINE:92316923; PubMed:1452287;
RA Zhou J., Hertz J.M., Leinonen A., Trygvasson K.;
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single base mutation in exon 28 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient";
RI J. Biol. Chem. 267:12475-12481(1992).
RN [3]
RP SEQUENCE OF 85-1685 FROM N.A.
RX TISSUE: Placenta;
RA MEDLINE:90337990; PubMed:2380186;
RA Pihlajaniemi T., Pihlajaniemi E.K., Myers J.C.;
RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5(IV)";
RI J. Biol. Chem. 265:13758-13766(1990).
RN [4]
RP SEQUENCE OF 924-1685 FROM N.A.
RX MEDLINE:91169491; PubMed:2004755;
RA Zhou J., Hostikka S.L., Chow L.T., Trygvasson K.;
RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome";
RI Genomics 9:1-9(1991).
RN [5]
RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE:90160375; PubMed:1689491;
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeythyaer M., Shown T.B., Trygvasson K.;
RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome";
RI Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
RN [6]
RP SEQUENCE OF 1442-1471 FROM N.A.

Query Match	7.98	Score	605.5	18	Length	1000
Best Local Similarity	24.28	Pred. NO.	1.90-15			
Matches 192	Conservative	96	Mismatches	542	Indels	94

40	PEYVWSTVSNQVLOJRSFNS	-----	OGKQJASASQSPVQJRSK	-----	406					
41										
199	PGGMLPGPPLVGLVGLKGMMLN	FOJRKGEGBDLOJPGPPLVGL	JSKQDITVEHVGK	2508						
42	69	-----	GCHPAAIPG	-----	410					
43										
259	DOGLDQDKRPPVGLKQPPPP	PGGEKFEKGEQEPKQKPOCKDQ	DNJQGLJGADIP	418						
44	111	RDQGTPNDSOTKFCNSADH	IKSDQSOHITPSKTPSMAFA	RSSTSHQJIALEPFAQ	1700					
45	319	GVYGEPOKQGE	-----	KQKQGTGPFG	PGVIAIP	-----	457			
46	171	KTFAVVVVVSTEVANKAAV	IKQVETIVSHIUNISNNK	IKRSTAPANTG	15ALBN	2008				
47	458	-----	KGN	-----	GLDLPGEKGEQ	-----	475			
48	241	KPLPPQPPAPANOJONSNT	RIQPTTPITAPAKPAAPR	PLDQBSVVENKL	11580	2908				
49	376	PGJGGTGP	-----	LPQPAAVVGGPVPQPPQKQK	QJHJH	PGJIS	417			
50	291	SH	-----	ASSTPLPACTQPNSTPN	RNAVTVSQSSSSAOPKAP	PEVSSDE	340			
51	418	LPQPLGLQJQVAPGLPQPP	QACHIPESDEICEP	-----	QPHQJNSQD	464				
52	441	PTLQENFDLNSQDLEHRE	SLSTLRJQMLFPDKEFT	IQVANSQDPOQSV	114P	498				
53	465	-----	KJQGGQGVAKKQKQD	ICEN	-----	GGJLSQSNQJQVLPJLQ	604			
54	399	QKKPPPLQAMMAOSQSL	AKQDQRTVQ	APQPPQJHIVPSP	-----	Q	444			
55	505	-----	PGSLG	PTQKQKQKQAGAL	QGRGLQIPQAVAPQPPQSKQJPD	5499				
56	444	EMVPPSNMSQSTQCPDH	LIHM	-----	TPQDIAMLKQGEF	FEKRRKQJVVVQVQ	510			
57	550	ILTFPMKQJGELQSDQAP	LIJLPGTQ	-----	Q	580				
58	499	DMVYQJHQRQV	-----	VKGPPYQOMPTSGMA	QKQTEPSSQJIMHSHLP	-----	516			
59	581	DOJLPQPPKQKPPQJLTK	QKQKPPG	-----	NPQLPLQNLNGP	-----	MGHGTQPPQJVS	622		
60	547	-----	KQMAPIHNMQCSQMLQ	QFAGM	INSEMGQNPVTPASQKQJAS	VSWPQ	-----	595		
61	633	EKGJQGVAGNQKQPG	-----	HPGPGHQQT	TOGRQDQSGNQKQDQMLPQ	QJH	QJH	688		
62	536	-----	VKJPIHQRFP	-----	PDQGTFSQNG	KG	EKFETPDQSLSEEMQDQJAFKQJAL	PIQMM	649	
63	689	PLDQCTFKSKQDQCP	PLQJLQAPPHQKQDQCTHQPQAP	-----	QJHJQJGL	745				
64	650	EMTRESMEKMMH	IKQSOHMEKQNNLP	-----	TRIP	VBQDQSPQKFEKQJLQ	700			
65	746	EIS	-----	PGQPPQPPQKQETQFAL	GGPQPPQJQKQALQ	KGQ	KQPPQ	782		
66	701	MGPPHQLPFGQVPSQKQDQV	LVNMSSQSNQ	POCKHEACQGEF	QKQLKQJASQSDQJLA	740				
67	783	PGPPQRTGLDQPL	GPQGV	-----	QPNQPPQM	-----	QPPQ	-----	PGJLVQGP	844
68	761	QOKWPLPQGE	-----	POQVYQMGPPPLPMSQ	PPG	SSNSQJLRLRLQJH	-----	807		
69	825	GPQCPQJQPPQJHGLPQKQ	QJDP	-----	PLQDVQPPQKQKQSPQ	PGQAVQJQJQJSGTLP	883			
70	808	-----	QRTNSQLSHM	PLQJLNFSSN	-----	TSJNAPVQKQJAKK	-----	844		
71	884	QKAGASQPPQKQKQMMQ	PPQPPQJLQJQBSQVQJL	KQJQDQJQVQPLPQPPQKQK	943					
72	845	-----	PLDJSVQSGVHS	-----	PGJNLKJSTJMQVQSP	PLQJASJNLKJQJTP	889			
73	944	QERPLVPPQPPQJNLQJQKQ	QKQPPQJL	-----	PGVSGKQJQVQJLQ	990				

```

UY      890  SOLQGMGLAGPAAASIIKSPVIVLGAASAAPVHLKSPSTLPATSPGMISSPEFELQSOGT - P 144
DB      961  GQ----- --POLSG- - - - - GPRLPGP-PGRKRNPPLDQ-PQLDP 162
UY      948  ENRKAVLT-TMASVAMLVNVSNGCPPTTA--SOPANVINIPLSL---PSSTLYIMPPED 999
DB      1024 FGLKGTIGMKGP-PGCG-VF--GGPSSVINGQSGMLGNQGRKGDYLSISLGFL 1077
UY      1000 TLNQNPILTIMMSKSKRAMSNITLYDALRYAASSLSINSHPAKRNILISMNNRPM-QI 1656
DB      1080 GRKCRP-----GLPGYGNPQIKGSVALFWALPW-LQGTHAKRGDGR-PGR 1122
UY      1059 -NTUMP----KISGP-NPVWPPETLSFGMGTPPL---SHSMNSGPNVWG----- 1099
DB      1124 PETGFHPGRKSLSGFPNGNLIP-----GGPVGSGGHGYPGPHGDKRKPDQDLTG 1177
UY      1100 -----FNIPHCVPMPGPMISHNPDMHQSSEPPVNVGMRMFP-----GGEPVV 1144
DB      1177 PAGCKGEIQGPFNGPNIHQFGLPETLSGGRKGKL--PGLP-GINDLPGRKHGPHRPVAV 124
UY      1145 QSPDVGVTFPIHNGSGGSG-SFGGGMFFPG-FGLPHSPNIDPOSSADALCKHGSGHY 1205
DB      1234 GQPP-----GGPSKGFALDSPKCNBPQQGPPGR-GLPQPE-----GPGQLG 1276
UY      1203 SEFTVGNMSPSVFTDPLDGLVTRNAGLPDEPLSLRTISEKISQTLQTFPRVEVGRKO 1266
DB      1277 NGGKRG-----FKONGQDGLP-----GLTKRG 1400
UY      1268 PQGP-----GGPSHMGMGMPLQAFHMTALPCMGP-----GWGTHDI-PLGFAF 1400
DB      1301 DQGPFIQGNIGRGRIKGMKRDPLKVPVRPMKRSVYGAGDGGRGLGPPG-PP 1455
UY      1309 SMPGHNRRPAPLVLGGCMGCEHHKMSPVAOSTMDPGPTLMSPNAACVMIPGRKSPAL 1468
DB      1360 GLPQSGG---SLIKGLDAP-----PGILGQGRKRG-----LPG-PGCVGL 1498
UY      1369 YTHGVPVSRG-----MMMSNGGEMGHTGQMIMIPGRHRGMADVOMGKPSGPNYS 1422
DB      1399 ---PGRTPGDPGRKNCLPGFTAGGRKGRGDLPGGPGIRGLDHPGGMASHGQPPGPG 1454
RESULT 2
CA24_ASCTU
ID CA24_ASCTU STANDARD; PERT: 1763 AA.
AC P2793;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 49, Last annotation update)
DB Collagen alpha 2(IV) chain precursor.
DC Collagen alpha 2(IV) chain precursor.
OS Ascaris suum (pig roundworm) (Ascaris lumbricoides).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OX NCBI_TaxId=6253;
   11
SP SEQUENCE FROM N.A.
RX MEDLINE-91340768; PubMed-1714907;
RA Pettit J., Kingston I.B.;
RT "The complete primary structure of a nematode alpha 2(IV) collagen
RT and the partial structural organization of its gene."
RL J. Biol. Chem. 266:16149-16156(1991).
CC -1- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
CC -1- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1
CC DOMAINS.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND IT: ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NON-COLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS. FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLETIDE REPEATING

```

UNIT (1 X 1) ARE HYDROLYZABLE IN SOME OR ALL OF THE OTHERS.
 1. TIME TYPE IV COLLATIONS CONTAIN INFORMATION SYSTEMS WITH THE
 ARE INVOLVED IN THE AND INFORMATIONAL INFORMATION IN THE
 THREE, LOCATED IN THE NOT MAIN, ARE CONSERVED IN ALL KNOWN TYPE
 IV COLLATIONS.
 2. THE SWISS PEOPLE IS COPYABLE. IT IS PRODUCED THROUGH A collaboration
 between the Swiss Institute of Hydroinformatics and the ETH Zurich
 the European hydroinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed, used by and for commercial
 or other purposes or for other purposes (see: <http://www.ethz.ch/infocenter/>
 or send an email to hydroinformatics@ethz.ch).
 3. 100% ME/100% AA/100% 1.1.
 4. 100% ME/100% ME/100% 1.1.
 5. 100% ME/100% ME/100% 1.1.
 6. 100% ME/100% ME/100% 1.1.
 7. 100% ME/100% ME/100% 1.1.
 8. 100% ME/100% ME/100% 1.1.
 9. 100% ME/100% ME/100% 1.1.
 10. 100% ME/100% ME/100% 1.1.
 11. 100% ME/100% ME/100% 1.1.
 12. 100% ME/100% ME/100% 1.1.
 13. 100% ME/100% ME/100% 1.1.
 14. 100% ME/100% ME/100% 1.1.
 15. 100% ME/100% ME/100% 1.1.
 16. 100% ME/100% ME/100% 1.1.
 17. 100% ME/100% ME/100% 1.1.
 18. 100% ME/100% ME/100% 1.1.
 19. 100% ME/100% ME/100% 1.1.
 20. 100% ME/100% ME/100% 1.1.
 21. 100% ME/100% ME/100% 1.1.
 22. 100% ME/100% ME/100% 1.1.
 23. 100% ME/100% ME/100% 1.1.
 24. 100% ME/100% ME/100% 1.1.
 25. 100% ME/100% ME/100% 1.1.
 26. 100% ME/100% ME/100% 1.1.
 27. 100% ME/100% ME/100% 1.1.
 28. 100% ME/100% ME/100% 1.1.
 29. 100% ME/100% ME/100% 1.1.
 30. 100% ME/100% ME/100% 1.1.
 31. 100% ME/100% ME/100% 1.1.
 32. 100% ME/100% ME/100% 1.1.
 33. 100% ME/100% ME/100% 1.1.
 34. 100% ME/100% ME/100% 1.1.
 35. 100% ME/100% ME/100% 1.1.
 36. 100% ME/100% ME/100% 1.1.
 37. 100% ME/100% ME/100% 1.1.
 38. 100% ME/100% ME/100% 1.1.
 39. 100% ME/100% ME/100% 1.1.
 40. 100% ME/100% ME/100% 1.1.
 41. 100% ME/100% ME/100% 1.1.
 42. 100% ME/100% ME/100% 1.1.
 43. 100% ME/100% ME/100% 1.1.
 44. 100% ME/100% ME/100% 1.1.
 45. 100% ME/100% ME/100% 1.1.
 46. 100% ME/100% ME/100% 1.1.
 47. 100% ME/100% ME/100% 1.1.
 48. 100% ME/100% ME/100% 1.1.
 49. 100% ME/100% ME/100% 1.1.
 50. 100% ME/100% ME/100% 1.1.
 51. 100% ME/100% ME/100% 1.1.
 52. 100% ME/100% ME/100% 1.1.
 53. 100% ME/100% ME/100% 1.1.
 54. 100% ME/100% ME/100% 1.1.
 55. 100% ME/100% ME/100% 1.1.
 56. 100% ME/100% ME/100% 1.1.
 57. 100% ME/100% ME/100% 1.1.
 58. 100% ME/100% ME/100% 1.1.
 59. 100% ME/100% ME/100% 1.1.
 60. 100% ME/100% ME/100% 1.1.
 61. 100% ME/100% ME/100% 1.1.
 62. 100% ME/100% ME/100% 1.1.
 63. 100% ME/100% ME/100% 1.1.
 64. 100% ME/100% ME/100% 1.1.
 65. 100% ME/100% ME/100% 1.1.
 66. 100% ME/100% ME/100% 1.1.
 67. 100% ME/100% ME/100% 1.1.
 68. 100% ME/100% ME/100% 1.1.
 69. 100% ME/100% ME/100% 1.1.
 70. 100% ME/100% ME/100% 1.1.
 71. 100% ME/100% ME/100% 1.1.
 72. 100% ME/100% ME/100% 1.1.
 73. 100% ME/100% ME/100% 1.1.
 74. 100% ME/100% ME/100% 1.1.
 75. 100% ME/100% ME/100% 1.1.
 76. 100% ME/100% ME/100% 1.1.
 77. 100% ME/100% ME/100% 1.1.
 78. 100% ME/100% ME/100% 1.1.
 79. 100% ME/100% ME/100% 1.1.
 80. 100% ME/100% ME/100% 1.1.
 81. 100% ME/100% ME/100% 1.1.
 82. 100% ME/100% ME/100% 1.1.
 83. 100% ME/100% ME/100% 1.1.
 84. 100% ME/100% ME/100% 1.1.
 85. 100% ME/100% ME/100% 1.1.
 86. 100% ME/100% ME/100% 1.1.
 87. 100% ME/100% ME/100% 1.1.
 88. 100% ME/100% ME/100% 1.1.
 89. 100% ME/100% ME/100% 1.1.
 90. 100% ME/100% ME/100% 1.1.
 91. 100% ME/100% ME/100% 1.1.
 92. 100% ME/100% ME/100% 1.1.
 93. 100% ME/100% ME/100% 1.1.
 94. 100% ME/100% ME/100% 1.1.
 95. 100% ME/100% ME/100% 1.1.
 96. 100% ME/100% ME/100% 1.1.
 97. 100% ME/100% ME/100% 1.1.
 98. 100% ME/100% ME/100% 1.1.
 99. 100% ME/100% ME/100% 1.1.
 100. 100% ME/100% ME/100% 1.1.

101. 100% ME/100% ME/100% 1.1.
 102. 100% ME/100% ME/100% 1.1.
 103. 100% ME/100% ME/100% 1.1.
 104. 100% ME/100% ME/100% 1.1.
 105. 100% ME/100% ME/100% 1.1.
 106. 100% ME/100% ME/100% 1.1.
 107. 100% ME/100% ME/100% 1.1.
 108. 100% ME/100% ME/100% 1.1.
 109. 100% ME/100% ME/100% 1.1.
 110. 100% ME/100% ME/100% 1.1.
 111. 100% ME/100% ME/100% 1.1.
 112. 100% ME/100% ME/100% 1.1.
 113. 100% ME/100% ME/100% 1.1.
 114. 100% ME/100% ME/100% 1.1.
 115. 100% ME/100% ME/100% 1.1.
 116. 100% ME/100% ME/100% 1.1.
 117. 100% ME/100% ME/100% 1.1.
 118. 100% ME/100% ME/100% 1.1.
 119. 100% ME/100% ME/100% 1.1.
 120. 100% ME/100% ME/100% 1.1.
 121. 100% ME/100% ME/100% 1.1.
 122. 100% ME/100% ME/100% 1.1.
 123. 100% ME/100% ME/100% 1.1.
 124. 100% ME/100% ME/100% 1.1.
 125. 100% ME/100% ME/100% 1.1.
 126. 100% ME/100% ME/100% 1.1.
 127. 100% ME/100% ME/100% 1.1.
 128. 100% ME/100% ME/100% 1.1.
 129. 100% ME/100% ME/100% 1.1.
 130. 100% ME/100% ME/100% 1.1.
 131. 100% ME/100% ME/100% 1.1.
 132. 100% ME/100% ME/100% 1.1.
 133. 100% ME/100% ME/100% 1.1.
 134. 100% ME/100% ME/100% 1.1.
 135. 100% ME/100% ME/100% 1.1.
 136. 100% ME/100% ME/100% 1.1.
 137. 100% ME/100% ME/100% 1.1.
 138. 100% ME/100% ME/100% 1.1.
 139. 100% ME/100% ME/100% 1.1.
 140. 100% ME/100% ME/100% 1.1.
 141. 100% ME/100% ME/100% 1.1.
 142. 100% ME/100% ME/100% 1.1.
 143. 100% ME/100% ME/100% 1.1.
 144. 100% ME/100% ME/100% 1.1.
 145. 100% ME/100% ME/100% 1.1.
 146. 100% ME/100% ME/100% 1.1.
 147. 100% ME/100% ME/100% 1.1.
 148. 100% ME/100% ME/100% 1.1.
 149. 100% ME/100% ME/100% 1.1.
 150. 100% ME/100% ME/100% 1.1.
 151. 100% ME/100% ME/100% 1.1.
 152. 100% ME/100% ME/100% 1.1.
 153. 100% ME/100% ME/100% 1.1.
 154. 100% ME/100% ME/100% 1.1.
 155. 100% ME/100% ME/100% 1.1.
 156. 100% ME/100% ME/100% 1.1.
 157. 100% ME/100% ME/100% 1.1.
 158. 100% ME/100% ME/100% 1.1.
 159. 100% ME/100% ME/100% 1.1.
 160. 100% ME/100% ME/100% 1.1.
 161. 100% ME/100% ME/100% 1.1.
 162. 100% ME/100% ME/100% 1.1.
 163. 100% ME/100% ME/100% 1.1.
 164. 100% ME/100% ME/100% 1.1.
 165. 100% ME/100% ME/100% 1.1.
 166. 100% ME/100% ME/100% 1.1.
 167. 100% ME/100% ME/100% 1.1.
 168. 100% ME/100% ME/100% 1.1.
 169. 100% ME/100% ME/100% 1.1.
 170. 100% ME/100% ME/100% 1.1.
 171. 100% ME/100% ME/100% 1.1.
 172. 100% ME/100% ME/100% 1.1.
 173. 100% ME/100% ME/100% 1.1.
 174. 100% ME/100% ME/100% 1.1.
 175. 100% ME/100% ME/100% 1.1.
 176. 100% ME/100% ME/100% 1.1.
 177. 100% ME/100% ME/100% 1.1.
 178. 100% ME/100% ME/100% 1.1.
 179. 100% ME/100% ME/100% 1.1.
 180. 100% ME/100% ME/100% 1.1.
 181. 100% ME/100% ME/100% 1.1.
 182. 100% ME/100% ME/100% 1.1.
 183. 100% ME/100% ME/100% 1.1.
 184. 100% ME/100% ME/100% 1.1.
 185. 100% ME/100% ME/100% 1.1.
 186. 100% ME/100% ME/100% 1.1.
 187. 100% ME/100% ME/100% 1.1.
 188. 100% ME/100% ME/100% 1.1.
 189. 100% ME/100% ME/100% 1.1.
 190. 100% ME/100% ME/100% 1.1.
 191. 100% ME/100% ME/100% 1.1.
 192. 100% ME/100% ME/100% 1.1.
 193. 100% ME/100% ME/100% 1.1.
 194. 100% ME/100% ME/100% 1.1.
 195. 100% ME/100% ME/100% 1.1.
 196. 100% ME/100% ME/100% 1.1.
 197. 100% ME/100% ME/100% 1.1.
 198. 100% ME/100% ME/100% 1.1.
 199. 100% ME/100% ME/100% 1.1.
 200. 100% ME/100% ME/100% 1.1.

DB 1415 GIPGIPAKGEGISLIGK RGNIGICKGPAIPILPMKKGSGHJGHPAAMJNL 1374
 UY 1398 PGMREPR---GMAAVCMKFSQSPGPNB 1422
 DB 1374 POLKGEPLPPPPGKGFCTGFCGCPGPG 1402

RESULT 3
 CA14_MOUSE STANDARD: PRT: 1669 AA.
 AC P02463;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 GN COLA1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX MGI_FaxID:10090;
 RN 11
 RP SEQUENCE FROM N.A.
 RA MEDLINE=89197932; PubMed=2703490;
 RT Muthukumar G., Blumberg B., Kurkinen M.;
 RT "The complete primary structure for the alpha 1-chain of mouse
 RT collagen IV. Differential evolution of collagen IV domains.";
 RL J. Biol. Chem. 264:6310-6317(1989).
 RN 12
 RP SEQUENCE OF 1-1154 FROM N.A.
 RA MEDLINE=88112221; PubMed=338568;
 RT Wood L., Thelander N., Vogel J.;
 RT "cDNA clones completing the nucleotide and derived amino acid
 RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
 RT from mouse.";
 RL FEBS Lett. 227:5-8(1988).
 RN 13
 RP SEQUENCE OF 1149-1424 FROM N.A.
 RA MEDLINE=86301886; PubMed=375692;
 RT Nath P., Laurent M., Horn E., Sobol M.E., Zon G., Vogel J.;
 RT "Isolation of an alpha 1 type IV collagen cDNA clone using a
 RT synthetic oligonucleotide.";
 RL Gene 43:301-304(1986).
 RN 14
 RP SEQUENCE OF 1276-1669 FROM N.A.
 RA MEDLINE=85127038; PubMed=2578961;
 RT Oberbauer L., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
 RT Vogel J., Voss T., Siebold R., Glauville R.W., Kuhn K.;
 RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
 RT the alpha 1(IV) chain of basement membrane collagen as derived from
 RT complementary DNA.";
 RL Eur. J. Biochem. 147:217-224(1985).
 RN 15
 RP SEQUENCE OF 1441-1669 FROM N.A.
 RA MEDLINE=87250440; PubMed=359738;
 RT Kurkinen M., Condon M.R., Blumberg B., Barlow D., Gullerens S.,
 RT Saus J., Pihlajaniemi T.;
 RT "Extensive homology between the carboxyl-terminal peptides of mouse
 RT alpha 1(IV) and alpha 2(IV) collagen.";
 RL J. Biol. Chem. 262:8496-8499(1987).
 RN 16
 RP PARTIAL SEQUENCE FROM N.A.
 RA MEDLINE=86196099; PubMed=3009408;
 RT Sakurai Y., Sullivan M., Yamada Y.;
 RT "Alpha 1 type IV collagen gene evolved differently from the alpha
 RT collagen genes.";
 RL J. Biol. Chem. 261:6654-6657(1986).
 RN 17
 RP SEQUENCE OF 1-28 FROM N.A.
 RA MEDLINE=89066738; PubMed=3198620;
 RT Kayes P., Wood L., Thelander N., Kurkinen M., Vogel J.;
 RT "Head-to-head arrangement of murine type IV collagen genes.";
 RL J. Biol. Chem. 263:19274-19277(1988).
 RN 18

RP SEQUENCE OF 1-28 FROM N.A.
 RA MEDLINE=89071759; PubMed=3200851;
 RT Bartholo P.D., Martin G.R., Yamada Y.;
 RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
 RT bidirectional promoter and a shared enhancer.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
 RN 19
 RP SEQUENCE OF 1-129 FROM N.A.
 RA MEDLINE=88243724; PubMed=339041;
 RT Kurkinen M., Bartholo P., Sakurai Y., Yamada Y.;
 RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
 RT collagen chain and the corresponding region of the gene.";
 RL J. Biol. Chem. 263:8706-8709(1988).
 CC 1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A "CHECKER BOARD"
 CC NETWORK TOGETHER WITH LAMININS, PROTEOLYNS AND ENKATIN/
 CC NIDOGEN.
 CC 1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOPFORMS, ALPHA 1(IV),
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC 1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS. FREQUENT INTERROPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL, TRIPLE HELICAL DOMAIN (WHICH
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC 1- PTM: PROTEINS AT THE THIRD POSITION OF THE TRIPLETTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC 1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: J03758; AAA37449.1; -
 DR EMBL: M23333; AAA51625.1; -
 DR EMBL: J04694; AAA50292.1; -
 DR EMBL: X06777; CAA29946.1; -
 DR EMBL: X02201; CAA26192.1; -
 DR EMBL: M15812; AAA37340.1; -
 DR EMBL: M14042; AAA37342.1; -
 DR EMBL: M12879; AAA37343.1; -
 DR EMBL: M13024; -; NOT_ANNOTATED_GDS.
 DR EMBL: M13025; -; NOT_ANNOTATED_GDS.
 DR EMBL: M13026; AAA37344.1; -
 DR EMBL: M13027; AAA37345.1; -
 DR EMBL: M13043; AAA37346.1; -
 DR EMBL: J04448; AAA37487.1; -
 DR PIR: A33525; GCM548.
 DR MGI: BB454; Col4a1.
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR001442; Procollagen*4.
 DR Pfam: PF01491; Collagen; 23.
 DR Pfam: PF01413; C4; 2.
 DR Pfam: PF000007; Collagen; 4.
 DR ProDom: PD003923; Procollagen*4; 2.
 DR SMART: SM00111; C4; 2.
 DR Extracellular matrix; Connective tissue; Basement membrane;
 DR Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
 KW SIGNAL.
 FT PROPEP 1 27
 FT CHAIN 173 1669 AMINO TERMINAL PROPEPTIDE (7S DOMAIN).
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
 FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
 FT DISULFID 1460 1561 OR 1548 (BY SIMILARITY).
 FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
 FT DISULFID 1505 1511 BY SIMILARITY.

[illegible][illegible]

KI Human tissues.
 KI J. Biol. Chem. 269:23013-23017(1994).
 KI [2]
 KI REVISIONS.
 KI Leinonen A.;
 KI Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 KI [3]
 KI SEQUENCE FROM N.A., VARIANTS AS E-297; R-407; R-640; R-1177; E-1207;
 KI U-1215; S-1277; T-1330; E-1344; E-1347 AND C-1661, AND VARIANTS R-433;
 KI R-162; Y-126; H-408; R-451; L-574; E-1269 AND P-1474.
 KI MEDLINE-21064696; PubMed 11134255;
 KI Heidet L., Arrondeau C., Forcster L., Cohen-Solal L., Muller G.,
 KI Gattelier B., Stavrou C., Gubler M.C., Auriama C.;
 KI "Structure of the human type IV collagen gene COL4A3 and mutations in
 KI autosomal Alport syndrome."
 KI J. Am. Soc. Nephrol. 12:97 106(2001).
 KI [4]
 KI SEQUENCE OF 1486-1670 FROM N.A., AND PARTIAL SEQUENCE.
 KI MEDLINE-93015826; PubMed-1400291;
 KI Quiñones S., Berнал D., García-Soto M., Elena S.F., Saus J.;
 KI "Exon/intron structure of the human alpha 3(IV) gene encompassing the
 KI Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
 KI antigenic region at the triple helix/NC1 domain junction."
 KI J. Biol. Chem. 267:19780-19784(1992).
 KI [5]
 KI SEQUENCE OF 1453-1670 FROM N.A.
 KI MEDLINE 91354570; PubMed-1882840;
 KI Morrison K.E., Matiyama M., Yang-Feng J.L., Reeders S.L.;
 KI "Sequence and localization of a partial cDNA encoding the human alpha
 KI 3 chain of type IV collagen."
 KI Am. J. Hum. Genet. 49:545-554(1991).
 KI [6]
 KI SEQUENCE OF 1331-1670 FROM N.A.
 KI TISSUE-Kidney;
 KI MEDLINE-92147878; PubMed 1737849;
 KI Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.;
 KI Pusey C.D.;
 KI "Molecular cloning of the human Goodpasture antigen demonstrates it
 KI to be the alpha 3 chain of type IV collagen."
 KI J. Clin. Invest. 89:592-601(1992).
 KI [7]
 KI SEQUENCE OF 1644-1670 FROM N.A.
 KI TISSUE-Kidney;
 KI Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
 KI [8]
 KI SEQUENCE OF 1449-1670, AND ALTERNATIVE SPLICING.
 KI TISSUE-Kidney;
 KI MEDLINE-94124597; PubMed 8294492;
 KI Feng L., Xia Y., Wilson C.H.;
 KI "Alternative splicing of the NC1 domain of the human alpha 3(IV)
 KI collagen gene: differential expression of mRNA transcripts that
 KI predict three protein variants with distinct carboxyl regions."
 KI J. Biol. Chem. 269:2342-2348(1994).
 KI [9]
 KI SEQUENCE OF 1-29 FROM N.A.
 KI MEDLINE-98196854; PubMed 9537506;
 KI Momota R., Sugimoto M., Gohashi T., Kikawa K., Yoshida H.,
 KI Niimura Y.;
 KI "Two genes, COL4A3 and COL4A3 coding for the human alpha3(IV) and
 KI alpha4(IV) collagen chains are arranged head-to-head on chromosome
 KI 2q36."
 KI FEBS Lett. 424:11-16(1998).
 KI [10]
 KI ALTERNATIVE SPLICING.
 KI MEDLINE-94280184; PubMed-8505332;
 KI Berнал D., Quiñones S., Saus J.;
 KI "The human mRNA encoding the Goodpasture antigen is alternatively
 KI spliced."
 KI J. Biol. Chem. 268:12090-12094(1993).
 KI [11]
 KI VARIANT PRO-1474.
 KI MEDLINE-95078827; PubMed 7987301;

RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
 RA Bartelds A., Momeni L.A.H., van Oost B.A., Reumer H.G.;
 RA Reenders S.T., Smeets H.J.M.;
 RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal
 RT recessive Alport syndrome."
 RT Hum. Mol. Genet. 3:1269-1273(1994).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLomerular basement membranes (GBM), FORMING A "CHICKEN-WIRE"
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOLYNS AND ENLACTIN/
 CC NIDOGIN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOBFORMS, ALPHA 1(IV)
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS: 1 (SHOWN HERE), 2/V AND
 CC 3/15; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR
 CC C-TERMINAL NC1 DOMAINS.
 CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
 CC COLocalIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
 CC COCHLEA, LUNG AND BRAIN.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERROGATIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL, TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLEPEPIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL
 CC PTM-TYPE IV COLLAGEN SITE.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -1- PTM: PHOSPHORYLATED BY THE Goodpasture antigen-binding protein.
 CC -1- DISEASE: ANTIBODIES AGAINST THE NC1 DOMAIN OF ALPHA3(IV) MEDIATE
 CC THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS
 CC CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.
 CC -1- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE 1
 CC AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY
 CC GLomerulonephropathy CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,
 CC HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN
 CC MALES AND FEMALES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X60031; CAA56335.1; -;
 DR EMBL: A2284487; CAC36101.1; JOINED.
 DR EMBL: A2284488; CAC36101.1; JOINED.
 DR EMBL: A2284489; CAC36101.1; JOINED.
 DR EMBL: A2284490; CAC36101.1; JOINED.
 DR EMBL: A2284491; CAC36101.1; JOINED.
 DR EMBL: A2284492; CAC36101.1; JOINED.
 DR EMBL: A2284493; CAC36101.1; JOINED.
 DR EMBL: A2284494; CAC36101.1; JOINED.
 DR EMBL: A2284495; CAC36101.1; JOINED.
 DR EMBL: A2284496; CAC36101.1; JOINED.
 DR EMBL: A2284497; CAC36101.1; JOINED.
 DR EMBL: A2284498; CAC36101.1; JOINED.
 DR EMBL: A2284499; CAC36101.1; JOINED.
 DR EMBL: A2285000; CAC36101.1; JOINED.
 DR EMBL: A2285001; CAC36101.1; JOINED.
 DR EMBL: A2285002; CAC36101.1; JOINED.
 DR EMBL: A2285003; CAC36101.1; JOINED.
 DR EMBL: A2285004; CAC36101.1; JOINED.
 DR EMBL: A2285005; CAC36101.1; JOINED.
 DR EMBL: A2285006; CAC36101.1; JOINED.
 DR EMBL: A2285007; CAC36101.1; JOINED.

[illegible]

07	50.2	WHEAT	RECEIVED	---	PLANT	---	SEWAGE	WTE	PERMANENT	54.0
08	0.9	REPAIRING	ANALYST	REPAIRING	WV	WV	WV	WV	WV	41.6
09	6.1	POSTAL	PROGRAM	WV	---	---	---	---	---	5.77
10	4.7	POST	---	---	---	---	---	---	---	4.73
11	5.76	---	---	---	---	---	---	---	---	6.19
12	4.74	LEAD	TO	REPAIRING	PLANT	---	---	---	---	5.26
13	6.29	REPAIRING	PLANT	---	---	---	---	---	---	6.59
14	6.00	WATER	REPAIRING	PLANT	---	---	---	---	---	6.43
15	6.44	REPAIRING	PLANT	---	---	---	---	---	---	7.12
16	7.14	REPAIRING	PLANT	---	---	---	---	---	---	7.03
17	7.14	REPAIRING	PLANT	---	---	---	---	---	---	7.72
18	7.04	REPAIRING	PLANT	---	---	---	---	---	---	7.08
19	7.73	REPAIRING	PLANT	---	---	---	---	---	---	8.26
20	7.73	REPAIRING	PLANT	---	---	---	---	---	---	7.04
21	8.27	REPAIRING	PLANT	---	---	---	---	---	---	8.78
22	7.05	REPAIRING	PLANT	---	---	---	---	---	---	8.40
23	8.73	REPAIRING	PLANT	---	---	---	---	---	---	9.02
24	8.41	REPAIRING	PLANT	---	---	---	---	---	---	9.00
25	9.03	REPAIRING	PLANT	---	---	---	---	---	---	9.48
26	9.01	REPAIRING	PLANT	---	---	---	---	---	---	9.56
27	9.49	REPAIRING	PLANT	---	---	---	---	---	---	9.94
28	9.57	REPAIRING	PLANT	---	---	---	---	---	---	10.04
29	9.95	REPAIRING	PLANT	---	---	---	---	---	---	10.44
30	10.05	REPAIRING	PLANT	---	---	---	---	---	---	10.90
31	10.55	REPAIRING	PLANT	---	---	---	---	---	---	11.11
32	10.61	REPAIRING	PLANT	---	---	---	---	---	---	11.06
33	11.12	REPAIRING	PLANT	---	---	---	---	---	---	11.66
34	11.07	REPAIRING	PLANT	---	---	---	---	---	---	11.57
35	11.63	REPAIRING	PLANT	---	---	---	---	---	---	12.12
36	11.68	REPAIRING	PLANT	---	---	---	---	---	---	12.14
37	12.13	REPAIRING	PLANT	---	---	---	---	---	---	12.73
38	12.15	REPAIRING	PLANT	---	---	---	---	---	---	12.71
39	12.72	REPAIRING	PLANT	---	---	---	---	---	---	13.41
40	12.92	REPAIRING	PLANT	---	---	---	---	---	---	12.99
41	13.42	REPAIRING	PLANT	---	---	---	---	---	---	13.66

[illegible][illegible]

TL submitted (DEC-1977) to the EIR data bank.
 FN [5]
 RP SEQUENCE OF 399-727.
 RX MEDLINE=79000143; PubMed-687591;
 RA Seyer J.M., Kang A.H.;
 RT Covalent structure of collagen: amino acid sequence of type
 RI consecutive CMB peptides from type III collagen of human liver.*;
 RL Biochemistry 17:404-341(1978).
 RN [6]
 RP SEQUENCE OF 728-964.
 RX MEDLINE=80198282; PubMed-6246925;
 RA Seyer J.M., Mainardi C., Kang A.H.;
 RT Covalent structure of collagen: amino acid sequence of alpha 1
 RI (111)-CH5 from type III collagen of human liver.*;
 RL Biochemistry 19:1583-1589(1980).
 RN [7]
 RP SEQUENCE OF 950-1466 FROM N.A.
 RX MEDLINE=88189627; PubMed-4357782;
 RA Manoo B.S., Palalaiish R.;
 RT Human pro alpha 1(III) collagen: cDNA sequence for the 5' end.*;
 RL Nucleic Acids Res. 16:2337-2347(1988).
 RN [8]
 RP REVISION TO 1184.
 RX MEDLINE=89084346; PubMed-3211760;
 RA Molyneux K., Balalajish R.;
 RT Human type III collagen 'variant' is a cDNA cloning artefact.*;
 RL Nucleic Acids Res. 16:11833-11833(1988).
 RN [9]
 RP SEQUENCE OF 1065-1466 FROM N.A.
 RX MEDLINE=85087944; PubMed 6096827.

¹⁸ Molecular cloning and carboxyl-propeptide analysis of human type III RT revealed a novel mutation.

RA procollagen-1.
 RL Nucleic Acids Res. 12:9483-9494(1984).
 RN [10]
 RP SEQUENCE OF 945-1200.
 RX MEDLINE 81208139; PubMed 7016180;
 RA Seyer J.M., Kang A.H.;
 RL "Covalent structure of collagen: amino acid sequence of alpha
 1(III)-C99 from type III collagen of human liver."
 RL Biochemistry 20:2621-2627(1981).
 RN [11]
 RP SEQUENCE OF 1176-1466 FROM N.A.
 RX MEDLINE=85157600; PubMed 2579949;
 RA Chu M.-L., Wei D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
 RL "Isolation of cDNA and genomic clones encoding human pro-alpha 1
 (I1) collagen. Partial characterization of the 3' end region of the
 gene."
 RL J. Biol. Chem. 260:4357-4463(1985).
 RN [12]
 RP SEQUENCE OF 1161-1200 FROM N.A.
 RX MEDLINE=86187804; PubMed=3754462;
 RA Misulfin M., Dalgleish R., Kluge-Brockerman B., Reuward S.L.,
 RL Tolstoshev P., Brantly M., Crystal R.G.;
 RL "Human type III collagen gene expression is coordinately modulated
 with the type I collagen genes during fibroblast growth."
 RL Biochemistry 25:1408-1414(1986).
 RN [13]
 RP SEQUENCE OF 1-170 FROM N.A.
 RX TISSUE-PLACENT;
 RL MEDLINE=88304360; PubMed=3405774;
 RA Tomin D., Riera G., de Crombrughe B.;
 RL "Nucleotide sequence of a cDNA coding for the amino-terminal region
 of human propro alpha 1(III) collagen."
 RL Nucleic Acids Res. 16:7201-7201(1988).
 RN [14]
 RP SEQUENCE OF 1-176 FROM N.A.
 RX MEDLINE=89378752; PubMed=2777084;
 RA Reuson-Chanda V., Su M.W., Wei D., Chu M.-L., Ramirez F.;
 RL "Cloning and analysis of the 5' portion of the human type-III
 procollagen gene (COL3A1)."
 RL Gene 78:255-265(1989).
 RL [15]

DB 744 -----GEPGGGAKKVKPKIDKDPDTGTPD-PPGAGGDPKDGKDAWTLPELAD 792
 QY 916 ASPVHLKSPSLPAPSP--WTSNKRPLDPSGIPPPNKALILMASAMQINVSQRR 971
 DB 793 -----PRSGGGRKPTGTPGACPGAGGAGGAGGAGGAGGAGGAGGAGGAG 841
 QY 972 --PTTASGFA -----SVNIPGSLDSSPTPTMPTPTPTPTPTPTPTPTPTPT 1016
 DB 842 AGPGGSGPAPGPPGPGVKGKGGSPGPGAGGAGGAGGAGGAGGAGGAGGAGGAG 890
 QY 1017 AMPSITLTHALAKTAVSSDIDSPASPMIISNNNNNNNNNNNNNNNNNNNNNN 1173
 DB 891 -----GPPGPGSG--PGKIDPPGPMNTGAGVSGVSLKIDAG 927
 QY 1074 MPTLSPKMTQPLSHSNQMSPNVAVCPNPPHGVPMGPMIIMSNPIIMT-HGSGEPHVP 1132
 DB 928 QP-----GKSGSPGAGCP-----PGAGGFLGAGITGAGAGLAMPQMG 966
 QY 1133 GRMGFGGFPFVPSPPQVYPPHNGSG--GUGSPGQMGFGE-GPIGPPNIPVSSAD 1189
 DB 967 -----PRGSP-----GPGVKGSGKPGQANLSEKIPG PGLPPLAGT 1206
 QY 1190 AALCKRGGPGPSPFTVLGSMSPVETDP---DIOGVLRGATGIPEDLSHPISEKP 1245
 DB 1007 AG--EPGRDGNPS-----DGLPRDPSGKGDJGCHNSGAPG----- 1244
 QY 1246 SUTLYQFPGCEVGRKQPPG--GPGFSIMQMGPGVAPRNGALP--MGAGHVGTP 1300
 DB 1045 -----ANGLRPPGVVGR--AGKSGDGSFSPAGPAGPAGANSHKAGHVG 1309
 QY 1401 DPLAGT-----APSMGN--NPMR-----PAFLQCGMGMPPHMMSPVAGSLMG 1433
 DB 1090 --PRGKGGFTGEGANGIKKHGRPPNPGAKSPGAG--QUGALG-----SHGPAF-PR 1139
 QY 1344 VPTLMSPPAAVMIPKIKRGTAGLYTHGPGVSP--MMMSMGGMGPGQNTIMPPMKR 1403
 DB 1140 GPVSPSP-----PGKD-GTSG--HDPGTGPPG-----PK 1166
 QY 1404 GMADVMGFGSGGPNPG 1422
 DB 1167 GNGGKSGSPSPGHPGPG 1185

RESULT 8
 CAL4.HUMAN
 ID CAL4.HUMAN STANDARD: PRT: 1669 AA.
 AC P02462;
 DT 21-JUL-1986 (Rel. 01, Created)
 DI 01-FEB-1996 (Rel. 33, Last sequence update)
 DI 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 GN COL4A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euarchontomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 GX MGI_TaxID 9606;
 RN 111
 RP SEQUENCE FROM N.A.
 RX MEDLINE:89340433; PubMed:2701944;
 RA Solinen R., Huotari M., Ganguly A., Prockop D.J., Trygvason K.;
 RT "Structural organization of the gene for the alpha 1 chain of human
 RT type IV collagen.";
 RL J. Biol. Chem. 264:13565-13571(1989).
 RN 121
 RP SEQUENCE OF 46-1257 FROM N.A.
 RX TISSUE-placenta;
 RC MEDLINE:88083584; PubMed:3691802;
 RA Solinen R., Haka-Risku T., Prockop D.J., Trygvason K.;
 RT "Complete primary structure of the alpha 1 chain of human basement
 RT membrane (type IV) collagen.";
 RL FEBS Lett. 225:188-194(1987).
 RN 131

RP SEQUENCE OF 1-943 FROM N.A.
 RC TISSUE-placenta;
 RX MEDLINE:88029471; PubMed:3311751;
 RA Brackel D., Oberhauser T., Dietlmeier H., Rabel W., Glanville R.W.,
 RA Deutzmann R., Kuehn K.;
 RT "Completion of the amino acid sequence of the alpha 1 chain of human
 RT basement membrane collagen (type IV) reveals 21 non-triplet
 RT interruptions located within the collagenous domain.";
 RL Eur. J. Biochem. 168:529-536(1987).
 RN 141
 RP SEQUENCE OF 28-243.
 RX MEDLINE:86004708; PubMed:4043082;
 RA Glanville R.W., Qian R.Q., Siebold R., Risteli J., Kuehn K.;
 RT "Amino acid sequence of the N-terminal degradation and cross-linking
 RT region (7S domain) of the alpha 1(IV) chain of human basement
 RT membrane collagen.";
 RL Eur. J. Biochem. 152:213-219(1985).
 RN 151
 RP SEQUENCE OF 534-1447.
 RX MEDLINE:85003629; PubMed:6434307;
 RA Rabel W., Glanville R.W.;
 RT "Structure of human basement membrane (type IV) collagen. Complete
 RT amino-acid sequence of a 914-residue-long pepsin fragment from the
 RT alpha 1(IV) chain.";
 RL Eur. J. Biochem. 143:545-556(1984).
 RN 161
 RP SEQUENCE OF 1256-1669 FROM N.A.
 RX MEDLINE:85207819; PubMed:2581969;
 RA Pihladantini T., Trygvason K., Myers J.C., Kirkman M., Lado R.,
 RA Cheung M.-C., Prockop D.J., Boyd C.D.;
 RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV
 RT procollagen reveal an unusual homology of amino acid sequences in two
 RT halves of the carboxyl-terminal domain.";
 RL J. Biol. Chem. 260:7681-7687(1985).
 RN 171
 RP SEQUENCE OF 1259-1669 FROM N.A.
 RX MEDLINE:85216555; PubMed:2582422;
 RA Brinker J.M., Guds L.J., Loid H.R., Ward S.-Y., Rosenblum J.,
 RA Kettlides N.A., Myers J.C.;
 RT "Restricted homology between human alpha 1 type IV and other
 RT procollagen chains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
 RN 181
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE:89034231; PubMed:3182844;
 RA Solinen R., Huotari M., Hostikka S.L., Prockop D.J., Trygvason K.;
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
 RT collagen are divergently encoded on opposite DNA strands and have an
 RT overlapping promoter region.";
 RL J. Biol. Chem. 263:17217-17220(1988).
 RN 191
 RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
 RX TISSUE-placenta;
 RC MEDLINE:89005112; PubMed:2844531;
 RA Siebold R., Deutzmann R., Kuehn K.;
 RT "The arrangement of intra- and intermolecular disulfide bonds in the
 RT carboxyl-terminal, non-collagenous aggregation and cross-linking domain
 RT of basement-membrane type IV collagen.";
 RL Eur. J. Biochem. 176:617-624(1988).
 CC 1-1. FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A "CHICKEN-WIRE"
 CC MESHWORK TOGETHER WITH LAMININS, PROTEGLYCAN AND ENACTIN/
 CC NIDOGIN.
 CC 1-1. SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV),
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC 1-1. DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NON-COLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC 1-1. PHE: LYSINES AT THE THIRD POSITION OF THE TRIPLE HELIX REPEATING
 CC UNIT ARE HYDROXYLATED IN ALL CASES AND BIND CARBOHYDRATE.

This Swiss patent entry is copyright ©. It is produced through a collaboration between the Swiss Institute of Biotechnology and the EBC, in connection with the European Biotechnology Institute. There are no restrictions on its use by non-profit institutions as long as its content is not used for any commercial and this statement is not removed, thereby by and for acknowledgment of our recognition of a license agreement (<http://www.eth.ch/biotechnology/>) of such content to license (eth.ch).

[illegible][illegible]

```

Db 640 K-I-V-P-L-P-G-P-A-F-G-L-K-G-S-P-A-----P-F-G-P-G-D-K-F-P-T-F-R-P-G-I-P-G-K-G-A-V-A-G-S-I-F-P 692
QY 594 D-V-F-K-P-I-P-K-N-P-P-P-O-A-T-I-S-G-P-K-E-R-E-P-N-P-O-L-S-E-E-M-Q-Q-O-A-E-R-Q-L-I-I-P-M-A-M-E-S 691
Db 693 G-P-P-G-K-P-V-I-G-L-P-Q-D-M-G-F-E-C-T-P-K-R-----P-G-P-N-I-P-S-N-P-V-Q-Q-K-G-E-H-V-I-L-I-K-I-L 746
QY 652 I-R-P-S-E-N-M-R-M-I-P-G-S-N-R-H-M-F-R-N-N-I-P-F--P-R-I-P-V-E-----G-P--E-S-I-S-R-D-F-P-K-I-P 698
Db 747 L-P-G-----P-G-T-----P-G-K-G-S-I-G-V-P-G-V-G-E-I-G-I-P-G-L-G-I-H-E-C-H-P-E-I-L-P-S-V-G 794
QY 699 -----P-O-M-G-P--G-R-E-L-E-F-G-M-V-P-S-M-K-G-V-N-I-V-N-M-G-----S-N-S-N-I-P-K-M 747
Db 795 S-P-G-V-G-I-G-P-P-A-R-G-P-P-G-Q-P-L-S-P-R-G-I-G-E-K-F-P-F-G-L-I-M-G-I-K-G-K-A-Q-U-I-P-I-I 854
QY 738 R-E-A-G-A-G-P-E-E-M-L-K-L-R-G-S-D-M-L-P-A-Q-Q-K-V-P-L-P-F-G-E-H-P-Q-E-Y-G-M-G-P-P-L-P-S-N-G-P-S-N-G-L 797
Db 855 G-Q-S-T-I-----P-G-----P-G-Q-G-A-G-I-P-G-F-P-S-K-G-E-M-V-M-T-F-P-Q-S-P-S-P-G-A-G-L 901
QY 798 K-N-L-K-P-I-G-P-P-K-I-N-S-K-I-S-H-M-P-L-I-N-I-S-S-N-T-I-S-I-N-I-A-P-V-Q-K-G-L-K-I-L-I-S-V-A-S-V-H-S 857
Db 902 -----P-G-K-G-D-----H-G-P-S-S-G-P-G-D-----G-L-K-G-K-G-I-W-G-L-I-V-----K 934
QY 858 P-G-----I-N-E-L-K-S-T-M-I-U-V-S-P-M-I-G--S-P-S-N-L-K-S-P-Q-T-E-S-Q-L-A--C-M-L-A-P-A-A-A-S-I-K-S 907
Db 935 P-G-S-M-K-V-I-M-O-S-M-K-O-K-G-O-G-E-K-A-Q-I-G-I-P-G-E-K-S-K-I-P-E-T-P-V-P-E-K-I-O-V-A-M-F-P-Q-P-K-G-D 994
QY 908 P-P-V-I-L-S-A-A-S-V-H-L-K-S-P-I-P-A-P--S-P-G-T-S-S-P-K-P-I-Q-S--G-I-P-I-P-E-N-K-A-P-I-T-M-A-S-P-M-G 964
Db 995 P-O-I-S-I-P-E-A-----P-E-L-P-I-P-K-S-V-A-C-M-G-I-G-T-F-G-E-K-G-V-I-T-R-I-Q-S-P-I-L-P-I-K-A-K-I 1047
QY 964 N-E-S-A-G-P-----P-P-I-N-S-O-V-A-S-V-I-N-I-S-I-S-I-Y-T-M-F-E-T-I-L 1001
Db 1048 E-K-Q-A-G-P-P-I-G-I-P-O-I-E-K-K-D-G-I-A-F-P-G-S-P-E-K-E-K-S-I-G-I-P-----M-E-N-S-T-L 1099
QY 1002 S-O-N-P-I-S-I-M-S-M-K-S-P-A-M-S-S-T-P-I-Y-H-A-I-K-T-V-A-S-I-D-S-P-A-R--S-P-N-L-S-M-M-M-M-M-I-N-I 1066
Db 1099 K-G-S-P-G-S-V-----G-T-G-S-G-L-----P-G-K-I-K-G-I-G-L-I-G-I-T-V-K-E-A 1145
QY 1061 Q-N-P-R-I-S-G-N-P-V-P-M-P-L-S-P-M-O-T-I-P-I-S-H-N-O-M-P-S---P-N-A-V-G---P-N-I-P-P-E-V-I-M-G-L 1114
Db 1146 G-I-P-T-P-G-----T-G-P-A-Q-K-G-E-P-S-I-G-I-G-S-A-N-G-E-R-K-I-L-P-K-E-C-H-G-H-A 1181
QY 1114 M-S-H-N-I-M-H-----G-S-Q-E-P-P-M-P--Q-G-N-M-G-F-----P-Q-G-P-P-V-S-P-Q-Q-V-P-P-H-N-P-S-A-Q 1162
Db 1182 K-G-D-K-S-K-E-F-V-G-P-P-L-A-G-S-P-G-I-P-G-S-G-E-G-F-M-G-P-P-Q-Q-P-I-P-S-I-G-A-T---E-R-P-K-I-R 1248
QY 1163 G-S-P-G-M-G-F-P-G-I-P-G-P-S-N-I-P-U-S-S-A-A-L-C-K-P-G-P-G-P-G-P-S-T-I-V-L-S-M-S-V-T-D-H-I-O 1221
Db 1239 G-P--G-Q-P-L-P-E-I-P-G-P-M-G--P-P-G-I-P--G-I-I-G-V-K-G-K-G-N-P-G-M-P-G-A-----P-G-V-G-P-K-G-D-I-P-E-V 1289
QY 1222 E-V-I-R-G-A-T-G-I-P-E-F-D-S-R-I-T-P-S-E-K-P-S-O-T-I-O-V-P-R-E-F-V-P-G-K-O-P-G-Q-G-P-E-S-I-M-O-M-M-G-P-O-A 1281
Db 1290 G-M--P-G-I-G-S-I-G-----I-T-S-K-G-I-M-G-----P-R-G--V-R-P-G-P-K-I-L--P-E-L-G-I-T-K-I-N-I-G-I-S 1345
QY 1282 P-R-M-G-I-A-I-P-G-M-G-P-S-V-G-T-P-D-I-P-I-A-T-S-M-P-G-H-N-I-M-R-R-P-A-I-Q-Q-S-M-O-M-I-H-R-M-S-P-A-S-I-T-M 1341
Db 1346 V-P-S-A-K-I-P--G-P-G-I-P-G-Y-D-I-I-K-D--F-P-Q-L-P-G--P-E-G-P-P-G-L---K-G-I-G-G 1376
QY 1342 P-Q-G-T-L-M-S-N-P-A-A-V-M-I-P-K-I-K-T-V-A-I-L-Y-H-G-V-G-S-P-R-M--M-S-M-G-M-M-I-O-Q-N-I-M-I-I-V 1399
Db 1377 P-G-----P-R-G-Q-G-V-T-G-I-A-V-I-G-P-G-I-P-G-D-A-P-Q-K-G-P-M-P-----A-G-P 1415
QY 1400 M-R-P-R-M-A-I-V-G-M-G-P-S-Q-G-P-N-I-G 1422
Db 1416 T-G-P-K-T-P-P-P-P-P-K-I-P-S-M-G-I-P-G 1438

```

```

DT 15-JUN-2002 (kol. 41, last annotation update)
DE Collagen alpha 1(VII) chain precursor (long-chain collagen) (1*
DE collagen).
CN COL7A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE 94327588; PubMed=8051117;
RA Christiano A.M., Greenspan D.S., Lee S., Uitto J.;
RT "Cloning of human type VII collagen, complete primary sequence of the
RL alpha 1(VII) chain and identification of intrachain polymorphisms."?
RN [2]
RP SEQUENCE OF 128-1493 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93384847; PubMed=1407247;
RA Christiano A.M., Roschbaum J.M., Chung-Hoiet L.C., Parente M.G.,
RA Woodley D.T., Pan T.C., Zhang R.Z., Chu M.-L., Hargreaves R.E.,
RA Uitto J.;
RT "The large non-collagenous domain (NC-1) of type VII collagen is
RT amino-terminal and chimeric. Homology to cartilage matrix protein,
RT the type III domain of fibronectin and the A domains of von
RL Wilbrand factor."?
RN [3]
RP Hum. Mol. Genet. 1:475-481(1992).
RX [4]
RP SEQUENCE OF 815-1419 FROM N.A.
RX MEDLINE=91334380; PubMed=1871109;
RA Parente M.G., Chung L.C., Kyriacou J., Woodley D.T., Wynn K.M.,
RA Bauer E.A., Martel M.-G., Chu M.-L., Uitto J.;
RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the
RL gene."?
RN [5]
RP Proc. Natl. Acad. Sci. U.S.A. 88:6941-6945(1991).
RX [6]
RP SEQUENCE OF 369-1255 FROM N.A.
RX MEDLINE=93107742; PubMed=1469284;
RA Gammon W.R., Abernethy M.L., Padilla K.M., Pitsayath P.S.,
RA Cook M.E., Wright J., Briggman R.A., Hunt S.W. III;
RT "Noncollagenous (NC1) domain of collagen VII resembles and binds to
RT adhesion proteins involved in tissue-specific organization of
RT extracellular matrix."?
RN [7]
RP J. Invest. Dermatol. 99:691-696(1992).
RX [8]
RP SEQUENCE OF 340-675 FROM N.A.
RP TISSUE=Keratinocytes;
RX MEDLINE=92231902; PubMed=1567409;
RA Tanaka T., Takahashi K., Furukawa F., Imamura S.;
RT "Molecular cloning and characterization of type VII collagen cDNA."?
RN [9]
RP Biochem. Biophys. Res. Commun. 183:958-961(1992).
RX [10]
RP SEQUENCE OF 2305-2944 FROM N.A.
RX MEDLINE=93271985; PubMed=8499916;
RA Greenspan D.S.;
RT "The carboxyl-terminal half of type VII collagen, including the non-
RT collagenous NC-2 domain and intron/exon organization of the
RT corresponding region of the COL7A1 gene."?
RN [11]
RP Hum. Mol. Genet. 2:273-278(1993).
RX [12]
RP SEQUENCE OF 1-87 FROM N.A.
RP TISSUE=Placenta;
RX MEDLINE=94375010; PubMed=8088784;
RA Christiano A.M., Notman G.G., Chung-Hoiet L.C., Lee S., Chenu M.,
RA Uitto J., Greenspan D.S.;
RT "Structural organization of the human type VII collagen gene (COL7A1),
RT composed of more exons than any previously characterized gene."?
RN [13]
RP Genomics 21:169-179(1994).
RX [14]
RP REVIEW ON GEB VARIANTS.
RX MEDLINE 98041696; PubMed 9475844;
RA Jorvikallio A., Pulkkinen L., Uitto J.;
RT "Molecular basis of dystrophic epidermolysis bullosa: mutations in
RT the type VII collagen gene (COL7A1)."?

```


[illegible]

OY	1252	PFGE-----	--VPRKQUD--	PG-----	-----
D6	2470	-PKEGKEPEPI REEDDERPODEHGLTGPFGSKRCKGACWNSAG KGRKGNANLIG	-----	-----	25
OY	1248	PGEFHMGMMGPGAPPMOCALPG-----MGGPGVATITDPL-LEIAPSMJHNMDP	-----	-----	1
RT	11				11
D6	2529	PCHPGACGMGEGRPR---GIIDMKPKPDNDGDVKRSKRPPDKISA GLIS-----	-----	-----	27
OY	1319	PAPIQGCMGPHHHMSIMOSTMPNPPILMSNP-----	-----	-----	14
D6	2579	----LKRLIABPGQ---PGAAHFGPR---GSFKDEVWGTHGKKELVGHPGSGKLGEH	-----	-----	2
OY	1453	---AAGVMJGK-DKGA-----GLYTHPGVGGSCM---MMSMDMMGNGONMLP	-----	-----	14
D6	2629	GVCACCLDEKDGKDPAGPSPGHPTAIHGKRGMGCPGVSTAMPKCKEHLGKRCRFR	-----	-----	26
OY	1498	POMRRGMAD-----VMGSPSQSGPHCN	1423	-----	1
D6	2689	GUPGPKMOGEGKCRHTPGIGFPRPSNMG	2720	-----	1
RESIST 10					
ID	CAZ4_CAEEL	STANDARD:	PR1: 1758 AA.		
DT	P17140: Q19098: Q19099:				
OT	01-AUG-1990 (Rel. 15, Created)				
HT	01-OCT-1994 (Rel. 30, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Collagen alpha 2(IV) chain precursor (lethal protein 2).				
GN	Lf1-2 OR CLF-1 OR FOLIG12-5.				
OS	Ctenorhabdilis elegans.				
QC	Eukaryotes; Metazoa; Nematoda; Chromadorea; Rhaditidae; Rhaditoidae;				
OC	Rhaditidae; Pelioderinae; Ctenorhabdilis.				
OX	NCBI_TaxID=6239;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND FUNCTION.				
KC	STRAIN=Bristol N2.				
FX	MEDLINE=94012964; PubMed 7691828;				
KA	Stidley M.H., Johnson J.D., Mellio G.C., Kramer J.M.:				
RT	"Genetic identification, sequencing, and alternative splicing of the				
RT	Ctenorhabdilis elegans alpha 2(IV) collagen gene."				
RL	J. Cell Biol. 123:255-264(1993).				
RN	[2]				
RP	PRELIMINARY SEQUENCE OF 1445-1758 FROM N.A.				
KC	STRAIN=Bristol N2.				
FX	MEDLINE=9008929; PubMed 2794871;				
KA	Gao X., Kramer J.M.:				
RT	"The two Ctenorhabdilis elegans basement membrane (type IV) collagen				
RT	genes are located on separate chromosomes."				
RL	J. Biol. Chem. 264:17574-17582(1989).				
RN	[3]				
RP	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.				
KC	STRAIN=Bristol N2.				
RA	Wu X., Le T.T.:				
RL	Submitted (Apr-1996) to the EMBL/Genbank/JDBI databases.				
RN	[4]				
RP	VARIANTS.				
FX	MEDLINE 94120591; PubMed 8045258;				
RA	Sidley M.H., Graham P.L., von Mendel N., Kramer J.M.:				
RT	"Mutations in the alpha 2(IV) basement membrane collagen gene of				
RT	Ctenorhabdilis elegans produce phenotypes of differing severities."				
RL	EMBO J. 13:38278-385(1994).				
CC	-1- FUNCTION: Collagen type IV is specific for basement membranes.				
CC	Vital for embryonic development.				
CC	-1- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.				
CC	TYPE IV COLLAGEN FORMS A MESH LIKE NETWORK LINKED THROUGH				
CC	INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1				
CC	DOMAINS.				
CC	-1- ALTERNATIVE PRODUCTS: 2 isoforms: 1/a (shown here) and 1/b; are				
CC	produced by alternative splicing.				
CC	-1- DEVELOPMENTAL STAGE: Isoform 1 is predominant in embryos and				
CC	isofom 11 is predominant in the larvae and adults.				

[illegible]

140	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----


```

UY 245 -----PQUTAPANODNSSNFTLQUTPPPA-----TAPKAAAPRQJORESPAVEN 284
DB 276 GEPUSPAMNAGOMBPGLPBERGRKQPSCTACARQNDKAAVAPVGLAPAG 341
UY 284 KLLPSVSNFASSTLPDUGTSPNSTPNNKRAVTPVSGSSNSSAPKRPPTVSS 344
DB 332 -----PPTPAAAG--AAGETTPUGARG-----SSTQSSKRPD--QPPGACAAAG 374
UY 344 LGENHKKLSQPLJHRBRSTQTLKDLTQRLPPEDEKPTQAQSSQVONVLLAP-- 399
DB 374 -AGNFGANGU-----PAKNTATGA-----PGLAATAPRFGA 404
UY 400 KKPEPTQAMMAOSOSLAGKQRTTWGAPFGVGHDPFSPUEWYPSMNSNGLICP 459
DB 404 KQPSGP-----QPSGAPGKGNNGEP--GAPGNKQDTAKDEPDAVQUPRPAQ 454
UY 460 DILDHMTPEOJAMLKQGEYEEERKQBOVVUQSTLOMMVYUHQHDEGVAKRTPPV 519
DB 454 -----BERKRI-ARDESPAG 468
UY 520 M--TPSGMAPOGTE--PESLAINMPSLP-----PRGMAPHNMWASJMG 566
DB 469 LPSGAPGKAPGSGRFGAINTAPKPPGEBKNSPVAVGKSGFGLAKRQ 527
UY 567 GMINSEKQNNVNPASRRLSVSWPDVYKTPHGNRPFGQITSGPG-RGR- 641
DB 528 GLTGS-----PGSRGPKRTSPPG-----PASQDQDHPAGP-----PWABQAVWGE 571
UY 642 PNVUGLSEFEMUQJAKKQJGTPPMMEGIRSMENRMP-----NSQHMEGNN 674
DB 572 PGRKGAAG--PGRKGRKAPGPG--AVGAKKQDEGACQPPPTGPAVGRKQVAV 628
UY 675 PTPRPLVPGPSPSRQDPFR-----GLPQMG--PGKELFGWYSSKQKOVNI 729
DB 529 PGRQGLP--GFAKPP--GEAKPQEGVVPAGAGP-----PAKGRKQTPGRKQV 677
UY 730 SOMIPKMRKAPGPPRMLKLKNGSDMLPAQKQWVLPGEHDPQYGN--GR- 766
DB 678 GPHQDQPGKANGAP-----GNKAKGACAGACAPGNGRPPPL 728
UY 787 MSGG-----PS-----SNSGRLKREPLGPORTNS--KLSHMPPLTNTSSNTSL 841
DB 729 GAKGDKRDPKCAKAPKQKQKGLTGLPGLPGLMADQKGEAPGPP--APPLGA 784
UY 832 NTAPVQKQJGKRELDISVAGSOVSPGINTPKSPIMHVOQSMLOSPSNLKSVP 891
DB 785 KMAPG--DRGPPGPPGAGAG-----EPGAD-----QPPAKRTEI 818
UY 892 LAGMLAGPAAASIKSPVIGSAASPVHLKSPSLPAPSWMTSSKPTLQSHGEPENH 951
DB 819 -----GDAKADACDP-----GPAQPTAPAPAGVAGAG--PKAGAGSAG 961
UY 952 APITMASPMAGVESGPPPTASPPASVNTGSLPSSVPTTMTPEPTLSQNTSTMS 1011
DB 862 -----ATGPPAAGRV--GPPGNS--GMLGPG-----PPGAPAKZ3 894
UY 1012 RMSKAMPSTPLYDAIKTVASSDDSPASRPNLPSSNNMGMKINTONPISSTNV 1071
DB 895 -----SKCP-----KEETPAGPAGPAGPAGP----- 918
UY 1072 VEMPTLSPMGMTPJLSSNGMSPNAVGNPPEHVPKMGJLMSNIPIMHJSSGPPMP 1131
DB 919 -----PQG-----EKGSPTAGP--TAPGTPOGUGIAGQGVATVNG 955
UY 1132 OGMHFTVGPFP--FVQSPVQVPPPHNGSGQGSF--PGMKRPPG--EGTGHPSNLQSS 1187
DB 956 KGRKGP--GLPQSGEKGU-----GPSASRSGRGPQPMGPGIAGPAG 1002
UY 1188 ADAALKPKGAPGAPGPGFTVJGNSMTSVTDLIDQVTRKAGATGTFEPLSKIT 1247
DB 1003 REGA--PQABAPAG-----RKAAG--PKQDQGTGTAQPPG-- 1035

```

```

UY 1248 TLQYFPGVYKQKQPGP--GFSIMQNMZQDAPRMLATPG--MCHQPVCTDPLAT 1405
DB 1036 -----ATG--ATGAPDPVPAKRNKRPRTFPAGPAGPAGPAGPAGPAG 1483
UY 1306 TA-----PSMPTNMTPTATFQCGMMGPHHKKNSIAQSTIMVQPLIMSNPAVGM 1457
DB 1084 DKGETDQMDRCKKRGCF-----SGLOG-----PQPRGAPGR 1125
UY 1458 LKCKRQGA-----CLYTHGCVSGNGMMSQNMZQVQUNIMTPQMRKMAAL 1408
DB 1126 -TAPGRFPAASACAKRKHNTNGJHGTGPG----- 1162
UY 1409 VEMGFSQSGPONG 1422
DB 1163 VQVPGPPGPPGPG 1176

RESULT 14
CA2B_MOUSE
ID CA2B_MOUSE STANDARD PRI: 1650 AA.
AC 064739; Q9Z1W0;
DT 15-JUN-1998 (rel. 36, created)
DT 40-MAY-2000 (rel. 39, last sequence update)
DT 15-JUN-2002 (rel. 41, last annotation update)
DE Collagen alpha 2(XI) chain precursor.
GN Col11A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId:10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:129/SvJ;
RA Rosen L., Qin S., Madan A., Loretz C., James R., Iorns M., Mix L.,
RA Hall J., Lasky S., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class II
RT region."
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RP [2]
RP SEQUENCE OF 1-1592 FROM N.A.
RC STRAIN:PV/N, and 129/SvJ TISSUE:Cartilage;
RC MEDLINE:97135795; PubMed:8981342;
RA Vandenberg P., Vuorio M., Ala-Kokko L., Prockop J.E.;
RT "The mouse col11a2 gene. Some transcripts from the adjacent ext-1a
RT gene extend into the col11a2 gene."
RL Mamm Biol. 15:359-367(1996).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY
CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS (BY SIMILARITY)
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 4(XI) IS A POST TRANSLATIONAL
CC MODIFICATION OF ALPHA 1(XI). ALPHA 4(XI) CAN ALSO BE FOUND INSTEAD
CC OF ALPHA 4(XI)-1(XI) (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -!- PIM: PROLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/annotation/
CC or send an email to license@ebi.ac.uk).
CC
DB EMBL: AF100956; AAC69905.1;
DB EMBL: U16789; AAA67751.1;
DB EMBL: U16790; AAA67752.1;
DB MGJ: M6188447; Col11A2.
DB InterPro: IPR000087; Collagen.
DB InterPro: IPR000885; Fib Collagen_C.
DB InterPro: IPR001791; Laminin_G.

```


- OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae;
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 1-472 FROM N.A.
 RX MEDLINE=89025644; PubMed=3178744;
 RA Tromp G., Kivimäki H., Stacey A., Shikata H., Balmain C.L.,
 RA Jantsch R., Prockop D.J.;
 RT "Structure of a full-length cDNA clone for the pro-alpha 1(I)
 RT chain of human type I procollagen.";
 RL Biochem. J. 253:919-922(1988).
 RN [2]
 RP SEQUENCE OF 1-181 FROM N.A.
 RX MEDLINE=84270697; PubMed=6462220;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morille M.,
 RA Myers J., Williams G., Ramirez F.;
 RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
 RT conservation of a pattern of introns and exons.";
 RL Nature 310:337-340(1984).
 RN [3]
 RP SEQUENCE OF 162-301.
 RX MEDLINE=71038625; PubMed=5529814;
 RA Click E.M., Bornstein P.;
 RT "Isolation and characterization of the cyanogen bromide peptides from
 RT the alpha 1 and alpha 2 chains of human skin collagen.";
 RL Biochemistry 9:4699-4706(1970).
 RN [4]
 RP SEQUENCE OF 263-268.
 RX MEDLINE=71001508; PubMed=4319110;
 RA Morgan P.H., Jacobs H.G., Stewart J.P., Cunningham L.W.;
 RT "A comparative study of glycopeptides from selected
 RT vertebrate collagens. A possible role of the carbohydrate in fibril
 RT formation.";
 RL J. Biol. Chem. 245:5042-5046(1970).
 RN [5]
 RP SEQUENCE OF 425-1464 FROM N.A.
 RX MEDLINE=84080385; PubMed=6689127;
 RA Bernard M.P., Chu M.-L., Myers J.G., Ramirez F., Eikenberry E.F.,
 RA Prockop D.J.;
 RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
 RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
 RT of structures that are conserved during evolution.";
 RL Biochemistry 22:5213-5221(1983).
 RN [6]
 RP SEQUENCE OF 1225-1454 FROM N.A.
 RX MEDLINE=88124208; PubMed=3340541;
 RA Maeckelae J.K., Raustina M., Virta A., Vuorio E.;
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 RT domain.";
 RL Nucleic Acids Res. 16:3349-3349(1988).
 RN [7]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=88097389; PubMed=3480510;
 RA Bornstein P., McKay J., Morishima J.K., Devaratnam S., Gellinas R.E.;
 RT "Regulatory elements in the first intron contribute to
 RT transcriptional control of the human alpha 1(I) collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8859-8873(1987).
 RN [8]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=8510970; PubMed=2857713;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
 RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
 RT Promoter structure, Alu repeats, and polymorphic transcripts.";
 RL J. Biol. Chem. 260:2315-2420(1985).
 RN [9]
 RP SEQUENCE OF 1-44 FROM N.A.
 RX MEDLINE=88033098; PubMed=2822714;
 RA Kossow C.M.S., Verrier W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
 RA de Wet W.J.;
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)
 RT collagen gene enhance transcription.";
 RL J. Biol. Chem. 262:15151-15157(1987).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91184577; PubMed=2810958;
 RA Kivimäki H., Tromp G., Prockop D.J.;
 RT "Mutations in collagen genes: causes of rare and some common diseases
 RT in humans.";
 RL FASEB J. 5:2052-2060(1991).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97259359; PubMed=9101290;
 RA Kivimäki H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibrillar
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [12]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91374476; PubMed=1895312;
 RA Byers P.H., Wallis G.A., Walling M.C.;
 RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
 RL J. Med. Genet. 28:433-442(1991).
 RN [13]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97169389; PubMed=9016532;
 RA Dajbajda R.;
 RT "The human type I collagen mutation database.";
 RL Nucleic Acids Res. 25:181-187(1997).
 RN [14]
 RP VARIANT OF-11 CYS-1166.
 RX MEDLINE=86287390; PubMed=3016737;
 RA Cole D.H., Byers P.H., Stelmann R.E.;
 RT "Lethal osteogenesis imperfecta resulting from a single nucleotide
 RT change in one human pro alpha 1(I) collagen allele.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
 RN [15]
 RP VARIANT OF-11 ARG-569.
 RX MEDLINE=87222295; PubMed=3108247;
 RA Bateman J.F., Chan D., Walters I.D., Rogers J.O., Cole W.C.;
 RT "Lethal perinatal osteogenesis imperfecta due to the substitution of
 RT arginine for glycine at residue 391 of the alpha 1(I) chain of type I
 RT collagen.";
 RL J. Biol. Chem. 262:7021-7027(1987).
 RN [16]
 RP VARIANT OF-11 CYS-926.
 RX MEDLINE=88033031; PubMed=3667599;
 RA Vogel R.E., Minor R.R., Froud M., Prockop D.J.;
 RT "A point mutation in a type I procollagen gene converts glycine 746
 RT of the alpha 1 chain to cysteine and destabilizes the triple helix in
 RT a lethal variant of osteogenesis imperfecta.";
 RL J. Biol. Chem. 262:14737-14744(1987).
 RN [17]
 RP VARIANT OF-11 ARG-842.
 RX MEDLINE=88249828; PubMed=3403550;
 RA Bateman J.F., Lammie S.R., Dahl H.H., Chan D., Cole W.C.;
 RT "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
 RT chain in lethal perinatal osteogenesis imperfecta: demonstration of
 RT the peptide defect by in vitro expression of the mutant cDNA.";
 RL J. Biol. Chem. 263:11627-11630(1988).
 RN [18]
 RP VARIANT OF CYS-1195.
 RX MEDLINE=89218628; PubMed=3243312;
 RA Labadie M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister L.W.;
 RT "A cysteine for glycine substitution at position 1017 in an alpha
 RT 1(I) chain of type I collagen in a patient with mild dominantly
 RT inherited osteogenesis imperfecta.";
 RL Mol. Biol. Med. 5:197-207(1988).
 RN [19]
 RP VARIANT OF-11 VAL-434.
 RX MEDLINE=89255493; PubMed=2470760;
 RA Patterson F., Smiley E., Bonadio J.;

61 "RNA sequence analysis of a putative fetal osteonectin repeat loca-
 62 tion." *J. Biol. Chem.* 264:10084-10087(1989).
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000

61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646

Fri Feb 28 16:45:29 2003

us-09-915-543-15.rsp

Page 29

100	1107	IKTHKGF-----SGLQDP----	PGPGSGSGEGEGSGSGSGAPGPGPGPGSGAGAGAKD	1154
QY	1363	KGFALYTHRPVGSPTGMMSSGGMGPGVONTMPTQMPKGMADVGRNATSGPNSG	1422	
Db	1154	-----GLNLPQTHSPG-----	-----PRKEDMAGVATPGPGPGPG	1107

Search completed: February 28, 2003, 16:05:51
Job time : 42 secs

[illegible][illegible]

[illegible][illegible]

[illegible]

ring-opening. A gamma-V configuration may be a functionalization of alpha (V) residues, a beta-oxo-epsilon-alanine (V) chain and one alpha (V) residue, initially linked by amide bonds, which the epsilon, is formed with disulfide cross-links made from lysine and a lysine residues.

A Model may place a restriction on the observed growth as well as on the error term. For instance, supportability may be a function of α (V) which, if the latter is taken to be terminal in the body of the knowledge, would yield a restriction that α is always less than or equal to the expected value of α (assuming statistical independence). This can be proved as follows:

[illegible][illegible]

2.7×10^{21} mol/L, 1.5×10^{21} mol/L, 1.0×10^{21} mol/L, 0.5×10^{21} mol/L, 0.25×10^{21} mol/L, 0.125×10^{21} mol/L, 0.0625×10^{21} mol/L, 0.03125×10^{21} mol/L, 0.015625×10^{21} mol/L, 0.0078125×10^{21} mol/L, $0.00390625 \times 10^{21}$ mol/L, $0.001953125 \times 10^{21}$ mol/L, $0.0009765625 \times 10^{21}$ mol/L, $0.00048828125 \times 10^{21}$ mol/L, $0.000244140625 \times 10^{21}$ mol/L, $0.0001220703125 \times 10^{21}$ mol/L, $0.00006103515625 \times 10^{21}$ mol/L, $0.000030517578125 \times 10^{21}$ mol/L, $0.0000152587890625 \times 10^{21}$ mol/L, $0.00000762939453125 \times 10^{21}$ mol/L, $0.000003814697265625 \times 10^{21}$ mol/L, $0.0000019073486328125 \times 10^{21}$ mol/L, $0.00000095367431640625 \times 10^{21}$ mol/L, $0.000000476837158203125 \times 10^{21}$ mol/L, $0.0000002384185791015625 \times 10^{21}$ mol/L, $0.00000011920928955078125 \times 10^{21}$ mol/L, $0.000000059604644775390625 \times 10^{21}$ mol/L, $0.0000000298023223876953125 \times 10^{21}$ mol/L, $0.00000001490116119384765625 \times 10^{21}$ mol/L, $0.000000007450580596923828125 \times 10^{21}$ mol/L, $0.0000000037252902984619140625 \times 10^{21}$ mol/L, $0.00000000186264514923095703125 \times 10^{21}$ mol/L, $0.000000000931322574615478515625 \times 10^{21}$ mol/L, $0.0000000004656612873077392578125 \times 10^{21}$ mol/L, $0.00000000023283064365386962890625 \times 10^{21}$ mol/L, $0.000000000116415321826934814453125 \times 10^{21}$ mol/L, $0.000000000058207660913467407171875 \times 10^{21}$ mol/L, $0.0000000000291038304567337035859375 \times 10^{21}$ mol/L, $0.00000000001455191522836685179296875 \times 10^{21}$ mol/L, $0.000000000007275957614183425896484375 \times 10^{21}$ mol/L, $0.0000000000036379788070917129482421875 \times 10^{21}$ mol/L, $0.000000000001818989403545856474121119375 \times 10^{21}$ mol/L, $0.000000000000909494701772928237060559375 \times 10^{21}$ mol/L, $0.0000000000004547473508864641185302796875 \times 10^{21}$ mol/L, $0.00000000000022737367544323205926513984375 \times 10^{21}$ mol/L, $0.000000000000113686837721616029632569921875 \times 10^{21}$ mol/L, $0.0000000000000568434188608080148162849609375 \times 10^{21}$ mol/L, $0.00000000000002842170943040400740814248046875 \times 10^{21}$ mol/L, $0.000000000000014210854715202003704071224034375 \times 10^{21}$ mol/L, $0.0000000000000071054273576010018520356120171875 \times 10^{21}$ mol/L, $0.00000000000000355271367880050092601780600859375 \times 10^{21}$ mol/L, $0.000000000000001776356839400250046308903004296875 \times 10^{21}$ mol/L, $0.0000000000000008881784197001250231544515021484375 \times 10^{21}$ mol/L, $0.00000000000000044408920985006251157722575101071875 \times 10^{21}$ mol/L, $0.000000000000000222044604925031255886112875505359375 \times 10^{21}$ mol/L, $0.0000000000000001110223024625156279430564377526796875 \times 10^{21}$ mol/L, $0.00000000000000005551115123125781397152821887633984375 \times 10^{21}$ mol/L, $0.000000000000000027755575615628906985764109438169921875 \times 10^{21}$ mol/L, $0.0000000000000000138777878078144534888820547190849609375 \times 10^{21}$ mol/L, $0.00000000000000000693889390390722674444102735954248046875 \times 10^{21}$ mol/L, $0.000000000000000003469446951953613372220513679771240234375 \times 10^{21}$ mol/L, $0.000000000000000001734723475976806686110256839885620119375 \times 10^{21}$ mol/L, $0.000000000000000000867361737988403344305528419944281059375 \times 10^{21}$ mol/L, $0.0000000000000000004336808689942016721527642099722140546875 \times 10^{21}$ mol/L, $0.0000000000000000002168404344971008360763821049861072234375 \times 10^{21}$ mol/L, $0.00000000000000000010842021724855041803819105249305361171875 \times 10^{21}$ mol/L, $0.00000000000000000005421010862427520901909552624652680589375 \times 10^{21}$ mol/L, $0.000000000000000000027105054312137604509547763123263402946875 \times 10^{21}$ mol/L, $0.0000000000000000000135525271560688022547738815611631712234375 \times 10^{21}$ mol/L, $0.00000000000000000000677626357803440112738694078058158561171875 \times 10^{21}$ mol/L, $0.00000000000000000000338813178901720056369347039029079280589375 \times 10^{21}$ mol/L, $0.000000000000000000001694065894508600281846735195145396402946875 \times 10^{21}$ mol/L, $0.000000000000000000000847032947254320140592367597572698201484375 \times 10^{21}$ mol/L, $0.0000000000000000000004235164736271600702961837987863491007234375 \times 10^{21}$ mol/L, $0.00000000000000000000021175823681358003514809189939317455036171875 \times 10^{21}$ mol/L, $0.00000000000000000000010587911840679001757404594969658727518089375 \times 10^{21}$ mol/L, $0.00000000000000000000005293955920339500878$

[illegible][illegible][illegible]

97	VYVLETHANAKAAAVLAKOVETIVSTHOMINENKATPEASADNINIGETALINIKOLIT	266
	- - - - -	
146	LIVVETHILL	285
	- - - - -	
204	EEVPLEDIOGULEVSTHBAANYEVEVSODTICAV	346
	- - - - -	

[illegible][illegible][illegible][illegible][illegible][illegible]

Case	Genotype	ESR (mm/sec)	Age
1	11111111	11	11
2	11111111	11	11
3	11111111	11	11
4	11111111	11	11
5	11111111	11	11
6	11111111	11	11
7	11111111	11	11
8	11111111	11	11
9	11111111	11	11
10	11111111	11	11
11	11111111	11	11
12	11111111	11	11
13	11111111	11	11
14	11111111	11	11
15	11111111	11	11
16	11111111	11	11
17	11111111	11	11
18	11111111	11	11
19	11111111	11	11
20	11111111	11	11
21	11111111	11	11
22	11111111	11	11
23	11111111	11	11
24	11111111	11	11
25	11111111	11	11
26	11111111	11	11
27	11111111	11	11
28	11111111	11	11
29	11111111	11	11
30	11111111	11	11
31	11111111	11	11
32	11111111	11	11
33	11111111	11	11
34	11111111	11	11
35	11111111	11	11
36	11111111	11	11
37	11111111	11	11
38	11111111	11	11
39	11111111	11	11
40	11111111	11	11
41	11111111	11	11
42	11111111	11	11
43	11111111	11	11
44	11111111	11	11
45	11111111	11	11
46	11111111	11	11
47	11111111	11	11
48	11111111	11	11
49	11111111	11	11
50	11111111	11	11
51	11111111	11	11
52	11111111	11	11
53	11111111	11	11
54	11111111	11	11
55	11111111	11	11
56	11111111	11	11
57	11111111	11	11
58	11111111	11	11
59	11111111	11	11
60	11111111	11	11
61	11111111	11	11
62	11111111	11	11
63	11111111	11	11
64	11111111	11	11
65	11111111	11	11
66	11111111	11	11
67	11111111	11	11
68	11111111	11	11
69	11111111	11	11
70	11111111	11	11
71	11111111	11	11
72	11111111	11	11
73	11111111	11	11
74	11111111	11	11
75	11111111	11	11
76	11111111	11	11
77	11111111	11	11
78	11111111	11	11
79	11111111	11	11
80	11111111	11	11
81	11111111	11	11
82	11111111	11	11
83	11111111	11	11
84	11111111	11	11
85	11111111	11	11
86	11111111	11	11
87	11111111	11	11
88	11111111	11	11
89	11111111	11	11
90	11111111	11	11

File	Size	MD5	SHA1	SHA256
6781-102801-102802-102803-102804-102805-102806-102807-102808-102809-102810-102811-102812-102813-102814-102815-102816-102817-102818-102819-102820-102821-102822-102823-102824-102825-102826-102827-102828-102829-102830-102831-102832-102833-102834-102835-102836-102837-102838-102839-102840-102841-102842-102843-102844-102845-102846-102847-102848-102849-102850-102851-102852-102853-102854-102855-102856-102857-102858-102859-102860-102861-102862-102863-102864-102865-102866-102867-102868-102869-102870-102871-102872-102873-102874-102875-102876-102877-102878-102879-102880-102881-102882-102883-102884-102885-102886-102887-102888-102889-102890-102891-102892-102893-102894-102895-102896-102897-102898-102899-102900-102901-102902-102903-102904-102905-102906-102907-102908-102909-102910-102911-102912-102913-102914-102915-102916-102917-102918-102919-102920-102921-102922-102923-102924-102925-102926-102927-102928-102929-102930-102931-102932-102933-102934-102935-102936-102937-102938-102939-102940-102941-102942-102943-102944-102945-102946-102947-102948-102949-102950-102951-102952-102953-102954-102955-102956-102957-102958-102959-102960-102961-102962-102963-102964-102965-102966-102967-102968-102969-102970-102971-102972-102973-102974-102975-102976-102977-102978-102979-102980-102981-102982-102983-102984-102985-102986-102987-102988-102989-102990-102991-102992-102993-102994-102995-102996-102997-102998-102999-103000-103001-103002-103003-103004-103005-103006-103007-103008-103009-103010-103011-103012-103013-103014-103015-103016-103017-103018-103019-103020-103021-103022-103023-103024-103025-103026-103027-103028-103029-103030-103031-103032-103033-103034-103035-103036-103037-103038-103039-103040-103041-103042-103043-103044-103045-103046-103047-103048-103049-103050-103051-103052-103053-103054-103055-103056-103057-103058-103059-103060-103061-103062-103063-103064-103065-103066-103067-103068-103069-103070-103071-103072-103073-103074-103075-103076-103077-103078-103079-103080-103081-103082-103083-103084-103085-103086-103087-103088-103089-103090-103091-103092-103093-103094-103095-103096-103097-103098-103099-103100-103101-103102-103103-103104-103105-103106-103107-103108-103109-103110-103111-103112-103113-103114-103115-103116-103117-103118-103119-103120-103121-103122-103123-103124-103125-103126-103127-103128-103129-103130-103131-103132-103133-103134-103135-103136-103137-103138-103139-103140-103141-103142-103143-103144-103145-103146-103147-103148-103149-103150-103151-103152-103153-103154-103155-103156-103157-103158-103159-103160-103161-103162-103163-103164-103165-103166-103167-103168-103169-103170-103171-103172-103173-103174-103175-103176-103177-103178-103179-103180-103181-103182-103183-103184-103185-103186-103187-103188-103189-103190-103191-103192-103193-103194-103195-103196-103197-103198-103199-103200-103201-103202-103203-103204-103205-103206-103207-103208-103209-103210-103211-103212-103213-103214-103215-103216-103217-103218-103219-103220-103221-103222-103223-103224-103225-103226-103227-103228-103229-103230-103231-103232-103233-103234-103235-103236-103237-103238-103239-103240-103241-103242-103243-103244-103245-103246-103247-103248-103249-103250-103251-103252-103253-103254-103255-103256-103257-103258-103259-103260-103261-103262-103263-103264-103265-103266-103267-103268-103269-103270-103271-103272-103273-103274-103275-103276-103277-103278-103279-103280-103281-103282-103283-103284-103285-103286-103287-103288-103289-103290-103291-103292-103293-103294-103295-103296-103297-103298-103299-103300-103301-103302-103303-103304-103305-103306-103307-103308-103309-103310-103311-103312-103313-103314-103315-103316-103317-103318-103319-103320-103321-103322-103323-103324-103325-103326-103327-103328-103329-103330-103331-103332-103333-103334-103335-103336-103337-103338-103339-103340-103341-103342-103343-103344-103345-103346-103347-103348-103349-103350-103351-103352-103353-103354-103355-103356-103357-103358-103359-103360-103361-103362-103363-103364-103365-103366-103367-103368-103369-103370-103371-103372-103373-103374-10337				

[illegible][illegible]

Year	Number of students	Number of teachers	Number of classes	Number of subjects	Number of hours
1994	100	10	10	10	1000
1995	110	11	11	11	1100
1996	120	12	12	12	1200
1997	130	13	13	13	1300
1998	140	14	14	14	1400
1999	150	15	15	15	1500
2000	160	16	16	16	1600
2001	170	17	17	17	1700
2002	180	18	18	18	1800
2003	190	19	19	19	1900
2004	200	20	20	20	2000
2005	210	21	21	21	2100
2006	220	22	22	22	2200
2007	230	23	23	23	2300
2008	240	24	24	24	2400
2009	250	25	25	25	2500
2010	260	26	26	26	2600
2011	270	27	27	27	2700
2012	280	28	28	28	2800
2013	290	29	29	29	2900
2014	300	30	30	30	3000
2015	310	31	31	31	3100
2016	320	32	32	32	3200
2017	330	33	33	33	3300
2018	340	34	34	34	3400
2019	350	35	35	35	3500
2020	360	36	36	36	3600
2021	370	37	37	37	3700
2022	380	38	38	38	3800
2023	390	39	39	39	3900
2024	400	40	40	40	4000
2025	410	41	41	41	4100
2026	420	42	42	42	4200
2027	430	43	43	43	4300
2028	440	44	44	44	4400
2029	450	45	45	45	4500
2030	460	46	46	46	4600
2031	470	47	47	47	4700
2032	480	48	48	48	4800
2033	490	49	49	49	4900
2034	500	50	50	50	5000
2035	510	51	51	51	5100
2036	520	52	52	52	5200
2037	530	53	53	53	5300
2038	540	54	54	54	5400
2039	550	55	55	55	5500
2040	560	56	56	56	5600
2041	570	57	57	57	5700
2042	580	58	58	58	5800
2043	590	59	59	59	5900
2044	600	60	60	60	6000
2045	610	61	61	61	6100
2046	620	62	62	62	6200
2047	630	63	63	63	6300
2048	640	64	64	64	6400
2049	650	65	65	65	6500
2050	660	66	66	66	6600
2051	670	67	67	67	6700
2052	680	68	68	68	6800
2053	690	69	69	69	6900
2054	700	70	70	70	7000
2055	710	71	71	71	7100
2056	720	72	72	72	7200
2057	730	73	73	73	7300
2058	740	74	74	74	7400
2059	750	75	75	75	7500
2060	760	76	76	76	7600

[illegible]

BY	802 - NIKSHULPSOLAM -	1ACAAAAAR	SEPVLAS 914
100	1	1	1
101	1	1	1
102	1	1	1
103	1	1	1
104	1	1	1
105	1	1	1
106	1	1	1
107	1	1	1
108	1	1	1
109	1	1	1
110	1	1	1
111	1	1	1
112	1	1	1
113	1	1	1
114	1	1	1
115	1	1	1
116	1	1	1
117	1	1	1
118	1	1	1
119	1	1	1
120	1	1	1
121	1	1	1
122	1	1	1
123	1	1	1
124	1	1	1
125	1	1	1
126	1	1	1
127	1	1	1
128	1	1	1
129	1	1	1
130	1	1	1
131	1	1	1
132	1	1	1
133	1	1	1
134	1	1	1
135	1	1	1
136	1	1	1
137	1	1	1
138	1	1	1
139	1	1	1
140	1	1	1
141	1	1	1
142	1	1	1
143	1	1	1
144	1	1	1
145	1	1	1
146	1	1	1
147	1	1	1
148	1	1	1
149	1	1	1
150	1	1	1
151	1	1	1
152	1	1	1
153	1	1	1
154	1	1	1
155	1	1	1
156	1	1	1
157	1	1	1
158	1	1	1
159	1	1	1
160	1	1	1
161	1	1	1
162	1	1	1
163	1	1	1
164	1	1	1
165	1	1	1
166	1	1	1
167	1	1	1
168	1	1	1
169	1	1	1
170	1	1	1
171	1	1	1
172	1	1	1
173	1	1	1
174	1	1	1
175	1	1	1
176	1	1	1
177	1	1	1
178	1	1	1
179	1	1	1
180	1	1	1
181	1	1	1
182	1	1	1
183	1	1	1
184	1	1	1
185	1	1	1
186	1	1	1
187	1	1	1
188	1	1	1
189	1	1	1
190	1	1	1
191	1	1	1
192	1	1	1
193	1	1	1
194	1	1	1
195	1	1	1
196	1	1	1
197	1	1	1
198	1	1	1
199	1	1	1
200	1	1	1

05	014 A AASHVAKSHAP	SHIVASHAKTIGUNOTYINIKALMASAM 901
	1 1 1 1	1 1 1 1
06	0055 AVALAKO1	PODUPPUPPUPPACAKADU EKUGUPPACIKOOL
		SHIVASHAKTA 114

[illegible][illegible]

OF 1104, BRYMIDOLAMINE, 10618, 10619, 10620, 10621, 10622, 10623, 10624, 10625, 10626, 10627, 10628, 10629, 10630, 10631, 10632, 10633, 10634, 10635, 10636, 10637, 10638, 10639, 10640, 10641, 10642, 10643, 10644, 10645, 10646, 10647, 10648, 10649, 10650, 10651, 10652, 10653, 10654, 10655, 10656, 10657, 10658, 10659, 10660, 10661, 10662, 10663, 10664, 10665, 10666, 10667, 10668, 10669, 10670, 10671, 10672, 10673, 10674, 10675, 10676, 10677, 10678, 10679, 10680, 10681, 10682, 10683, 10684, 10685, 10686, 10687, 10688, 10689, 10690, 10691, 10692, 10693, 10694, 10695, 10696, 10697, 10698, 10699, 10700, 10701, 10702, 10703, 10704, 10705, 10706, 10707, 10708, 10709, 10710, 10711, 10712, 10713, 10714, 10715, 10716, 10717, 10718, 10719, 10720, 10721, 10722, 10723, 10724, 10725, 10726, 10727, 10728, 10729, 10730, 10731, 10732, 10733, 10734, 10735, 10736, 10737, 10738, 10739, 10740, 10741, 10742, 10743, 10744, 10745, 10746, 10747, 10748, 10749, 10750, 10751, 10752, 10753, 10754, 10755, 10756, 10757, 10758, 10759, 10760, 10761, 10762, 10763, 10764, 10765, 10766, 10767, 10768, 10769, 10770, 10771, 10772, 10773, 10774, 10775, 10776, 10777, 10778, 10779, 10780, 10781, 10782, 10783, 10784, 10785, 10786, 10787, 10788, 10789, 10790, 10791, 10792, 10793, 10794, 10795, 10796, 10797, 10798, 10799, 10800, 10801, 10802, 10803, 10804, 10805, 10806, 10807, 10808, 10809, 10810, 10811, 10812, 10813, 10814, 10815, 10816, 10817, 10818, 10819, 10820, 10821, 10822, 10823, 10824, 10825, 10826, 10827, 10828, 10829, 10830, 10831, 10832, 10833, 10834, 10835, 10836, 10837, 10838, 10839, 10840, 10841, 10842, 10843, 10844, 10845, 10846, 10847, 10848, 10849, 10850, 10851, 10852, 10853, 10854, 10855, 10856, 10857, 10858, 10859, 10860, 10861, 10862, 10863, 10864, 10865, 10866, 10867, 10868, 10869, 10870, 10871, 10872, 10873, 10874, 10875, 10876, 10877, 10878, 10879, 10880, 10881, 10882, 10883, 10884, 10885, 10886, 10887, 10888, 10889, 10890, 10891, 10892, 10893, 10894, 10895, 10896, 10897, 10898, 10899, 10900, 10901, 10902, 10903, 10904, 10905, 10906, 10907, 10908, 10909, 10910, 10911, 10912, 10913, 10914, 10915, 10916, 10917, 10918, 10919, 10920, 10921, 10922, 10923, 10924, 10925, 10926, 10927, 10928, 10929, 10930, 10931, 10932, 10933, 10934, 10935, 10936, 10937, 10938, 10939, 10940, 10941, 10942, 10943, 10944, 10945, 10946, 10947, 10948, 10949, 10950, 10951, 10952, 10953, 10954, 10955, 10956, 10957, 10958, 10959, 10960, 10961, 10962, 10963, 10964, 10965, 10966, 10967, 10968, 10969, 10970, 10971, 10972, 10973, 10974, 10975, 10976, 10977, 10978, 10979, 10980, 10981, 10982, 10983, 10984, 10985, 10986, 10987, 10988, 10989, 10990, 10991, 10992, 10993, 10994, 10995, 10996, 10997, 10998, 10999, 11000, 11001, 11002, 11003, 11004, 11005, 11006, 11007, 11008, 11009, 11010, 11011, 11012, 11013, 11014, 11015, 11016, 11017, 11018, 11019, 11020, 11021, 11022, 11023, 11024, 11025, 11026, 11027, 11028, 11029, 11030, 11031, 11032, 11033, 11034, 11035, 11036, 11037, 11038, 11039, 11040, 11041, 11042, 11043, 11044, 11045, 11046, 11047, 11048, 11049, 11050, 11051, 11052, 11053, 11054, 11055, 11056, 11057, 11058, 11059, 11060, 11061, 11062, 11063, 11064, 11065, 11066, 11067, 11068, 11069, 11070, 11071, 11072, 11073, 11074, 11075, 11076, 11077, 11078, 11079, 11080, 11081, 11082, 11083, 11084, 11085, 11086, 11087, 11088, 11089, 11090, 11091, 11092, 11093, 11094, 11095, 11096, 11097, 11098, 11099, 11100, 11101, 11102, 11103, 11104, 11105, 11106, 11107, 11108, 11109, 11110, 11111, 11112, 11113, 11114, 11115, 11116, 11117, 11118, 11119, 11120, 11121, 11122, 11123, 11124, 11125, 11126, 11127, 11128, 11129, 11130, 11131, 11132, 11133, 11134, 11135, 11136, 11137, 11138, 11139, 11140, 11141, 11142, 11143, 11144, 11145, 11146, 11147, 11148, 11149, 11150, 11151, 11152, 11153, 11154, 11155, 11156, 11157, 11158, 11159, 11160, 11161, 11162, 11163, 11164, 11165, 11166, 11167, 11168, 11169, 11170, 11171, 11172, 11173, 11174, 11175, 11176, 11177, 11178, 11179, 11180, 11181, 11182, 11183, 11184, 11185, 11186, 11187, 11188, 11189, 11190, 11191, 11192, 11193, 11194, 11195, 11196, 11197, 11198, 11199, 1120

[illegible][illegible]

	THE	1986	ACADEMIC YEAR	FOR	ACHIEVEMENT IN THE WORLD'S BEST SCHOOLS	IN	SCIENCE
57	1-50	STANDARD-S&P	SNO	AAS&M	P&K	DIRECTORY OF VISIONARY SCHOOLS	10
	(1-1)	(1-1)		(1-1)	(1-1)	(1-1)	

[illegible][illegible]

[illegible][illegible]

[illegible][illegible]

QY 769 EGEHVEEYGM -GKRTL----- PMSGHNSNLSKRLPTLITWKR 813
 DB 644 CIGDPEENIKPGPGKGVGAGAHGKRDSDAPDERPPDACLPAKQIAITPPDRE 704
 QY 811 NSRLSHHPLPLNNSMDTSLNTAPVQRLGKPLDLSVAGSVHSHNINLKSLTMH 876
 DB 704 GKKPDAPIPHVANSND-GLQGMHGRGPG-----SKGKIK 742
 QY 871 VSHMLASPSG---NLKSLPTLSQALADLAPAAASLKSPYLASAASTVHLKSTSL 946
 DB 743 -----GEPDACAADGVCKMGRGAPACIGPPDPAQ---QDILKRDSD 784
 QY 927 P-----AVSPMTSSPKPLDQSVGIPPHKAVLIMASPMGLGVNVSAPHTFASQA 979
 DB 787 PGLAGPGHGEHGEHPPGPACTPMAMONGERAGKGVAKVEKGEHPPDIA 841
 QY 980 SVNTFSLSTPTMTPEPTLSQNPSTMSKSKAMSNSTLYHAKIVANSIDDS 1039
 DB 842 -----GPISSGP-AGPPD-----GVK---GEGGS 864
 QY 1040 P-PARSHPLSMNNPDMGINTQNPRTSGNPVPMPLSP-MGMTPDLSHNSQMSPN 1096
 DB 865 PCPPGTAGPGGKRLKGPNNNNP---GP-----PQSNAPGKRIHGPANSSPPNPG 917
 QY 1097 AVGP-----NIPP-HGVPMGPKLSHNPIMG-HGSGPPMVYGBMGPMPV 1144
 DB 918 IAGKGDAGGHEGKGPVAGSPGSPGLAGLTARGAGLGPDMG-----EGGSF 970
 QY 1145 QSPPOVPPHNPSS---GGSHPPGMPGPG-GLPLRPNLPLNSDAALDKRHHGHP 1201
 DB 971 -----GPGGKHSNKPANSHNGERPPG-PGCLPQVQITAG EGRKGNP 1015
 QY 1202 DSEPTALNSMVSPTLP---DLEVTIPGATGIEEDLSRLLSEKISLAEYEEGEV 1257
 DB 1016 GS---DQDGRGRSSNKKDRGNSSPCAPG-----A 1044
 QY 1258 PGRKQPPG-GIGFSHGMGMGHPMGLALP---MAGHGVHPPHPIA 1407
 DB 1045 PGRGPPVPPG-SKSGSGEGEGFAGPSADPAGAKAPGPGG---PGRKKEIGE 1098
 QY 1408 -----PSMGRHNPRTPATLQDGMGMHMHMMSPASPMKQVLTAMNDAAV 1455
 DB 1099 KQSNGLKGRHGFEGNPG-PPSSFGAACHQDAIS-----SSJ-----TADGR 1148
 QY 1456 GM-----PKDGRGACLTYPHPPVSPGRMMSSQGMGQUNIMLPTQMHPROMAIVG 1410
 DB 1149 GPVGHPPPKKD-GTSG---HPGPIGPPG-----PKNMGEGEG 1172
 QY 1411 MGRSGGPGNIG 1422
 DB 1173 SSGSGHGGGNG 1184
 RESULT 14
 A54649
 collagen alpha 1(VII) chain precursor human
 M:Alternate names: procollagen alpha 1(VII) chain
 C:Species: Homo sapiens (man)
 C:Date: 04-Nov-1994 #sequence revision 04-Nov-1994 #ext-change 20-Sep-1999
 C:Accession: A54649; PH0844; S16416; S16426; A10296; I64686
 R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Nitto, J.
 J. Biol. Chem. 269, 20256-20262, 1994
 A:Title: Cloning of human type VII collagen, complete primary structure of the alpha1(VII)
 A:Reference number: A54649; MIMD:94327568; PMID:8051117
 A:Accession: A54649
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-2944 -CHR-
 A:Cross-references: GB:102870; NID:9487124; P1IN:AAA9449.1; F1:0498725
 R:Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.
 Biochem. Biophys. Res. Commun. 185, 958-965, 1992
 A:Title: Molecular cloning and characterization of type VII collagen cDNA.

A:Reference number: PH0844; MIMD:92241902; PMID:1567409
 A:Accession: PH0844
 A:Molecule type: mRNA
 A:Residues: 7EPK, 440-475, KAISTASHSTLWKATKMHVNHKSHHFAAEVPPNKAASHBAVAAT, 524-528,
 A:Cross-references: DBJ:0111152; DBJ:0116943; NID:9454698; P1IN:AAA9485.1; F1:04546
 A:Experimental source: Keratinocyte
 A:Note: The authors translated the codon ATG for residues 394 and 497 as 15T
 R:Parante, M.G.; Chong, L.C.; Ryanen, J.; Woodley, D.L.; Wynn, K.C.; Bauer, E.A.;
 Proc. Natl. Acad. Sci. U.S.A. 88, 6941-6945, 1991
 A:Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
 A:Reference number: S16316; MIMD:91334380; PMID:1871109
 A:Accession: S16316
 A:Molecule type: mRNA
 A:Residues: 815-892, 7E, 894-1439 -PAR-
 A:Cross-references: GB:M65158; GB:S49017; NID:910914; P1IN:AAA9649.1; F1:0490915
 A:Experimental source: Keratinocyte
 R:Kamman, W.R.; Abnerthy, M.L.; Padilla, K.M.; Pilsbury, P.S.; Cook, M.F.; Wight, J.
 Invest. Dermatol. 99, 691-696, 1992
 A:Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion p
 A:Reference number: 156328; MIMD:94107742; PMID:1469284
 A:Accession: 156328
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 7EPK, 372-517, 7GV, 520-540, 7W, 542-1255 -KES-
 A:Cross-references: GB:S51235; NID:9262408; P1IN:AAH4467.1; F1:0426409
 R:Seitzler, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Ghanville, R.W.; Pearson, R
 J. Biol. Chem. 264, 3822-3826, 1989
 A:Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagen
 A:Reference number: A30296; MIMD:89138437; PMID:2547292
 A:Accession: A30296
 A:Molecule type: protein
 A:Residues: A, 1240-1246, 7G, 1248-1250, 7XP, 1253-1255, 7V, 1257, 7E, 1262, 7V, 2034-204
 A:Note: two reported peptides cannot be reliably located
 R:Greenpan, D.S.
 Hum. Mol. Genet. 2, 273-278, 1993
 A:Title: The carboxyl-terminal half of type VII collagen, including the non collageno
 A:Reference number: 148103; MIMD:93271985; PMID:849916
 A:Accession: 148103
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 2395-2871, 7S, 2874-2944 -RE2-
 A:Cross-references: GB:106862; NID:9388713; P1IN:AAA9196.1; F1:0488714
 R:Christiano, A.M.; Ryanen, M.; Nitto, J.
 Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3554, 1994
 A:Title: Dominant dystrophic epidermolysis bullosa: identification of a glycosyl
 A:Reference number: A55255; MIMD:94224777; PMID:8170945
 A:Contents: annotation
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating and
 ed and subsequently O-glycosylated.
 C:Genetics:
 A:Gene: COL7A1; EBR1; EBR1; EB
 A:Cross-references: GB:128750; GIMM:120120
 A:Map position: 3p21.3-3p21.3
 A:Note: defects in this gene can result in dominant and recessive dystrophic epidermo
 A:Note: there are 118 introns
 C:Complex: type VII collagen is probably a homotrimer
 C:Function:
 A:Description: structural component of extracellular polymer associated with anchori
 C:Significance: massed collagen; animal Kunitz type proteinase inhibitor homology
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxyllysine; hydroxypr
 F:1-1/Domain: signal sequence #status predicted -SIG-
 F:17-2944/Product: collagen alpha 1(VII) chain #status predicted -MAT-
 F:17-1254/Domain: amino-terminal noncollagenous #status predicted -N1-
 F:36-201/Domain: von Willebrand factor type A repeat homology -VWA1-
 F:231-316/Domain: fibronectin type III repeat homology -FN1-
 F:327-413/Domain: fibronectin type III repeat homology -FN2-
 F:414-502/Domain: fibronectin type III repeat homology -FN3-
 F:508-594/Domain: fibronectin type III repeat homology -FN4-
 F:598-684/Domain: fibronectin type III repeat homology -FN5-
 F:686-771/Domain: fibronectin type III repeat homology -FN6-
 F:776-862/Domain: fibronectin type III repeat homology -FN7-
 F:864-952/Domain: fibronectin type III repeat homology -FN8-
 F:954-1045/Domain: fibronectin type III repeat homology -FN9-

[illegible][illegible]

A: Reference number: 220611
 A: Accession: 129431
 A: Status: preliminary; translated from gb/EMBL/DBJ
 A: Molecule type: DNA
 A: Molecule type: DNA
 A: Residues: 1-1759 (57%)
 A: Cross-references: EMBL:05442; PDB:AAA9215.1; GSPDB:000028; EST:001012.5a
 A: Experimental source: strain Bristol N2; clone F01G12
 C: Genes:
 A: Gene: GSP: F01G12.5a
 A: Map position: X
 A: Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 265/1; 304/1; 359/1; 450/2; 737/1
 C: Superfamily: collagen alpha 1(I) chain
 Query Match 6.78; Score 516.5; DB 2; Length 1759;
 Best Local Similarity 23.84; Pred. No. 6.1e-15;
 Matches 356; Conservative 104; Mismatches 537; Indels 15; Gaps 82;

QY 238 PPAVANDUNSSONTROTPPTAPAP-----KIAAPPTLDRESVVENKLLP 287
 DB PGPFGNKGDM-----KPPPPPPAPGMDGYNQIDAGATPPKPPPPMPPPPPP 149
 QY 288 SV-----GSPASSTELPP-----DTGPNST---PNNRAVTPVSQ--SNSSAATKAPPP 333
 DB 150 GMDKAKPPATGAGAGGCKATGAGKMPGPPGPPGSSGKGTGKGGGGLGNAHPPPP 209
 QY 334 PVPSSGPPPTLGENPDGLSQGLEHRRSLQTLRDIGRMLPPDEKPTAGSSATGQNG 394
 DB 210 EAGSPGNP-----GSPGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 229
 QY 394 VLD--GTPKKPPPTGAMMAVSSSLG-----KG--PPRPPVAP----- 429
 DB 230 DTCAMP--APPPPPPLASTMSKGTLLPPKGLGKGGKGGPPPPPPPPPPPPPP 288
 QY 430 -----PPGQCHRVPPSPPEMVPSPNSNSGTLGPHHLMPPEDLAWKIQEVEY 480
 DB 289 PPKKGGKGLSTAPKPKKPKKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPP 342
 QY 481 EKKKKKPPVVGQSLQMMVHGHPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 543
 DB 343 PGR-----DPPKPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 369
 QY 541 PHSLP-----PKMAPHPPNMG-----SOMRPPGACMNSMPPPPNPPPP 594
 DB 370 PPGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 427
 QY 585 PGLSGVMPDVPKLPDGRPP-----PQD-----GTFSSPPK 637
 DB 428 PGARVVSQTSGLTGLDPPKPPPPKPPPPKPPPPKPPPPKPPPPKPPPPKPP 487
 QY 618 -----GRR--FTNPPGLSEEMPOQGLAEKQLDPPPMAMEGLK----- 664
 DB 488 GPPGDSLNTLPPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 540
 QY 655 SPMKMKKPPGSGHPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 697
 DB 541 GTPGKGLPPVSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 599
 QY 598 PPGMG-----PPRPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 721
 DB 600 PPKKPP 659
 QY 722 LNVNMSNSQ-----MPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPP 766
 DB 660 LPPKSKPP 712
 QY 767 LPPGHPPP 817
 DB 713 MPPGPP 772
 QY 818 PPLPP 875
 DB 773 PPLPP 819

QY 876 LQSHSNLKSPPTSPQACMAGHAAASLSPVLAASAAVHUKSTLAPAS 943
 DB 820 VQVP--GPPGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 870
 QY 934 TSSPPKPPQ-----SPGPPPPKPPPPPPPPPPPPPPPPPPPPPPPPPP 976
 DB 871 SPP 925
 QY 977 -----QPAVNPVPSLSSPTPTPTPTLSONPLSMSSKPPPPPPPPPP 1042
 DB 926 GIDGAPGPP 957
 QY 1033 ASSDDP-----SLPAPSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1077
 DB 958 PPKKPP 1017
 QY 1078 -----SPGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1127
 DB 1018 NGLPPKPP 1079
 QY 1128 PVPVPP 1173
 DB 1077 P-----GPKPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1130
 QY 1174 -----PP 1221
 DB 1131 PPKKPP 1184
 QY 1222 -----PP 1265
 DB 1185 KPP 1238
 QY 1266 PPKPP 1316
 DB 1239 PPKPP 1291
 QY 1317 PPKPP 1376
 DB 1291 PPKPP 1347
 QY 1377 PPKPP 1426
 DB 1348 PPKPP 1393

Search completed: February 28, 2003, 16:06:43
 Job time: 58 secs



[illegible][illegible]

```

07 115 TTNHSHDKRNSAHLKSSQSHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
08 279 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 11AHEH1AKK 172
09 173 FAKVYVSTHMANKA AFAVIKQVETVSHLONISNNKERSHAIUNLUSAN 1b 228
10 445 -----LADIMANOKKRAH146 ----- 11AHEH1AKK 172
07 229 NUPKHPDQPPAVANADONSSNONTALPPTTPHFAHAK ----- 11AHEH1AKK 172
10 465 TVQUTSHHPPIYPPISQQAASHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 281 VBNKLHPS -----VGSASS ----- 11AHEH1AKK 172
10 443 LNKSHASSPSSFOGSSASTYNOUOQMDPBPONNPIHQBQGVSSSHQBNHBYQ 402
07 416 -----PVSQSSSSSNAH1KAP ----- 11AHEH1AKK 172
10 443 GNVPIHFAVWQQPPHNOGSHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 450 GHSQDLEH -----EKSLOTLHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 543 NSGAVUQANQVQAGHAGGAGHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 389 QUNGVUHLHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 603 QVNSHNSQSHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 446 -----KQVHSHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 663 VNPISQSHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 488 QVNVVQSHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 721 MGNKQVHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 526 -----WAPVHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 781 MGLQVHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 547 -----KCM-----APHHPHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 841 ASGHSHSHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 569 -----TNSHSHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 901 GVNKKHSHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 597 PKHSHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 961 PKHSHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 648 EKQVHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1011 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 11AHEH1AKK 172
07 689 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1064 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 11AHEH1AKK 172
07 738 KFAVHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1119 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 11AHEH1AKK 172
07 771 EHVHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1179 AHPHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 840 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1231 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172

```

```

07 873 SPMLSPSHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1288 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 944 WSHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1446 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 985 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1488 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 1027 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1447 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 1074 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1507 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 1120 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1565 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 1180 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1602 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 1242 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1653 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 1264 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1713 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 1329 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1773 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 1468 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1828 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
07 1403 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172
10 1888 SHHPIHSHSTSSNAAPRSHSTSSBQ 11AHEH1AKK 172

```

```

RESULT 3
US-09-915-497-56
Sequence 56, Application US/09919497
Patent No. US2002010662A1
GENERAL INFORMATION:
APPLICANT: Mottet, George L.
TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOCRINE CANCER
FILE REFERENCE: H08017225
CURRENT APPLICATION NUMBER: US/099194,497
CURRENT FILING DATE: 2001-07-31
PRIORITY APPLICATION NUMBER: US 60/221,745
PRIORITY FILING DATE: 2000-07-31
NUMBER OF SEQ ID NOS: 100
SOFTWARE: Patent in version 4.0
SEQ ID NO 56
LENGTH: 1806
TYPE: PRP
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: UNSURE
LOCATION: (758)
OTHER INFORMATION: Xaa - any amino acid
NAME/KEY: UNSURE
LOCATION: (807)

```


Country	Match	Score	4-6-8	10-12	14-16	17-19	20-22	23-25	26-28	29-31	32-34	35-37	38-40	41-43	44-46	47-49	50-52	53-55	56-58	59-61	62-64	65-67	68-70	71-73	74-76	77-79	80-82	83-85	86-88	89-91	92-94	95-97	98-100	Total	Rank	Points	Games	Matches	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419	2420	2421	2422	2423	2424
---------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	-------	------	--------	-------	---------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

[illegible]

[illegible][illegible]

Database version 1.9.4
Copyright (c) 1993 - 2003 Computer Ltd.

4M protein protein search, using sw model

Run on: February 26, 2003, 16:04:09 : Search time: 21 seconds

(with 1000000 updates/sec)

Title: US-09-915-543-15

Percent score: 76.2

Sequence: 1 MSHMKVRSFSCNTQSP.....ALWMDQSGDQGNINMMF (127)

Scoring table:

Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262571

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post Processing: Minimum Match: 0%

Listing first 45 summaries

Database:

- 1: /seq2_5/p/oldata/1/seq/5A_000.ppt*
- 2: /seq2_5/p/oldata/1/seq/5B_000.ppt*
- 3: /seq2_5/p/oldata/1/seq/5C_000.ppt*
- 4: /seq2_5/p/oldata/1/seq/5D_000.ppt*
- 5: /seq2_5/p/oldata/1/seq/5E_000.ppt*
- 6: /seq2_5/p/oldata/1/seq/5F_000.ppt*

Prod. No. is the number of results predicted by choice of database
Score greater than or equal to the score of the result being printed
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Post-Filter
1	606	7.9	960	US-09-219-849-5	Sequence 5, App 11
2	565	7.3	1064	US-08-642-255-62	Sequence 62, App 1
3	549.5	7.0	897	US-08-347-633A-50	Sequence 52, App 1
4	528	6.9	1078	US-08-963-825-21	Sequence 21, App 1
5	528	6.9	1078	US-09-570-811-21	Sequence 21, App 1
6	528	6.9	1078	US-09-570-811-21	Sequence 21, App 1
7	528	6.9	1078	US-09-548-608-21	Sequence 21, App 1
8	524.5	6.8	1057	US-08-941-820-4	Sequence 42, App 1
9	515.5	6.7	847	US-08-175-15-68	Sequence 10, App 1
10	515.5	6.7	847	US-08-175-15-68	Sequence 10, App 1
11	515.5	6.7	847	US-08-175-15-68	Sequence 10, App 1
12	515.5	6.7	847	US-08-175-15-68	Sequence 10, App 1
13	515.5	6.7	847	US-08-175-15-68	Sequence 10, App 1
14	515.5	6.7	847	US-08-175-15-68	Sequence 10, App 1
15	515.5	6.7	847	US-08-175-15-68	Sequence 10, App 1
16	508.5	6.6	1418	US-09-219-849-5	Sequence 5, App 11
17	506	6.6	1461	US-09-585-887-9	Sequence 9, App 1
18	506	6.6	1461	US-09-585-887-9	Sequence 9, App 1
19	504.5	6.6	1418	US-08-963-825-21	Sequence 21, App 1
20	504.5	6.6	1418	US-09-570-811-21	Sequence 21, App 1
21	504.5	6.6	1418	US-09-570-811-21	Sequence 21, App 1
22	504.5	6.6	1418	US-09-548-608-21	Sequence 21, App 1
23	502.5	6.5	1442	US-08-941-820-4	Sequence 42, App 1
24	502.5	6.5	1442	US-08-941-820-4	Sequence 42, App 1
25	500	6.5	1060	US-08-941-820-4	Sequence 42, App 1
26	497.5	6.5	1694	US-08-941-820-4	Sequence 42, App 1
27	493.5	6.4	1057	US-08-941-820-4	Sequence 42, App 1

28	493.5	6.4	1441	US-08-963-825-21	Sequence 21, App 1
29	493.5	6.4	1441	US-09-570-811-21	Sequence 21, App 1
30	493.5	6.4	1441	US-09-570-811-21	Sequence 21, App 1
31	493.5	6.4	1441	US-09-548-608-21	Sequence 21, App 1
32	488	6.3	829	US-08-642-255-62	Sequence 62, App 1
33	488	6.3	829	US-08-642-255-62	Sequence 62, App 1
34	478	6.2	1065	US-08-642-255-62	Sequence 62, App 1
35	478	6.2	1065	US-08-642-255-62	Sequence 62, App 1
36	443.5	5.8	777	US-08-642-255-62	Sequence 62, App 1
37	439	5.7	1466	US-09-219-849-5	Sequence 5, App 11
38	439	5.7	1466	US-09-219-849-5	Sequence 5, App 11
39	439	5.7	1466	US-09-585-887-9	Sequence 9, App 1
40	439	5.7	1466	US-09-585-887-9	Sequence 9, App 1
41	434.5	5.6	1024	US-08-941-820-4	Sequence 42, App 1
42	434	5.6	1466	US-09-585-887-9	Sequence 9, App 1
43	434	5.6	1466	US-09-585-887-9	Sequence 9, App 1
44	412.5	5.4	761	US-08-707-247A-84	Sequence 84, App 1
45	409	5.3	762	US-08-642-255-62	Sequence 62, App 1

ALIGNMENTS

RESULT 1

US-09-219-849-5

Sequence 5, Application US-09-219-849

Patent No. 6150061

GENERAL INFORMATION:

APPLICANT: VAN HERDE, GEORGE V.

APPLICANT: VAN KLIN, ALEXIS C.

APPLICANT: HOMSTRA, JAN B.

APPLICANT: DE WOLF, FREDERIK A.

APPLICANT: MEERBROEK, ANDREAS

APPLICANT: WERLEN, MARK W.L.

APPLICANT: WIND, RICHARD D.

APPLICANT: VAN DEN BROEK, TANJA J.

TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH PROMINANT POLYMER

TITLE OF INVENTION: SOLUBLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE

TITLE OF INVENTION: PREPARATION METHOD

FILE REFERENCE: 2728 2

CURRENT FILING DATE: 1998-12-23

CURRENT FILING DATE: 1998-12-23

NUMBER OF SEQ ID NOS: 50

SOFTWARE: Patent in Ver. 2.1

SEQ ID NO: 5

LENGTH: 960

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Illustrative

OTHER INFORMATION: amino acid sequence

US-09-219-849-5

Query Match

Best Local Similarity: 26.2% Prod. No. 4, 60 seqs

Matches: 155; Conserved: 47; Mismatches: 499; Indels: 510; Gaps: 79

UY	54	ASASQSLSTSSGSHHFKALPGVSSMFKKANKAGAKRKKPSISASPSFHH 114	114
UY	1	CAHCHACHKACHGAG--PKGAGPVAHPSKRGHAGV--FAHPSHFG 49	49
UY	114	GTPTNSDTEKNSADTKRSQSHHPSMFAHPSHSDGDAHFAHFK 114	114
UY	50	GPV-----AGVAFHPSHFG 74	74
UY	174	AKVYVSTEMANKAAVAKGVETLVSEHNTSSNKKPSFATNTLSAIKFKP 244	244
UY	75	A-----APPSK 86	86
UY	244	LPQGPANANGQSSSNTKLPPLPPAPAKPKAAHPPHDKSVKNTLVSVSVA 294	294
UY	86	---PAGVAGPVSADP-----PPHPSADPAGPVSADP-----FAHPSHFG 110	110

[illegible]

UY 562 -----LPGPAJ-----MINSEMERN-----VFNVANSNLSVSWITGV 596
 1b 323 PGDIPANLPLAABRKAJLSRHPAJOJGJHGEKPAFKARPAJAGKADAEJBDJAV 482
 UY 597 PKIPGJNREHUGGJFSGJSGEJEPJNPGJSEJPPJQJAEKJGLJPSJMAJHJHJGSM 656
 1b 483 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 449
 UY 657 ERMJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 715
 1b 441 ---MGPJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 471
 UY 715 MGPJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 772
 1b 472 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 520
 UY 773 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 842
 1b 521 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 775
 UY 843 JAPVJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 888
 1b 579 GMPJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 942
 UY 886 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 949
 1b 613 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 1000
 UY 940 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 982
 1b 661 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 725
 UY 983 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 1042
 1b 721 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 747
 UY 1043 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 1099
 1b 748 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 792
 UY 1100 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 1158
 1b 793 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 826
 UY 1159 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 1215
 1b 827 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 879
 UY 1216 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 1272
 1b 879 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 910
 UY 1271 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 1313
 1b 911 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 966
 UY 1314 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 1369
 1b 967 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 1008
 UY 1370 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 1426
 1b 1009 PGJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJHJGJH 1056

1 TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 2 TITLE OF INVENTION: In Body Fluids; A Test Kit and Means for Carrying out the
 3 TITLE OF INVENTION: Method and Use of the Method for Diagnosing the Presence of
 4 TITLE OF INVENTION: Disorders Associated with the Metabolism of
 5 NUMBER OF SEQUENCE: 21
 6 CORRESPONDENCE ADDRESS:
 7 ADDRESSEE: Doty & Doty PC
 8 STREET: 605 Third Avenue
 9 CITY: New York
 10 STATE: New York
 11 COUNTRY: USA
 12 ZIP: 10022
 13 COMPUTER READABLE FORM:
 14 METHOD TYPE: Floppy disk
 15 COMPUTER: IBM PC compatible
 16 OPERATING SYSTEM: PC-DOS/MS-DOS
 17 SOFTWARE: Patent Release #1.0, Version #1.25
 18 CURRENT APPLICATION DATA:
 19 APPLICATION NUMBER: US-09/500,811
 20 FILING DATE:
 21 CLASSIFICATION:
 22 PRIOR APPLICATION DATA:
 23 APPLICATION NUMBER: 08/187,419
 24 FILING DATE:
 25 ATTORNEY/AGENT INFORMATION:
 26 NAME: Gabor G. Ador
 27 REGISTRATION NUMBER: 29,714
 28 REFERENCE/WORK NUMBER: 4405/08701
 29 TELECOMMUNICATION INFORMATION:
 30 TELEPHONE: 212 527-7700
 31 TELEFAX: 212 753 6237
 32 TELEX: 246687
 33 INFORMATION FOR SEQ ID NO: 21:
 34 SEQUENCE CHARACTERISTICS:
 35 LENGTH: 1078 amino acids
 36 TYPE: amino acid
 37 TOPOLOGY: linear
 38 MOLECULE TYPE: protein
 39 ORIGINAL SOURCE:
 40 ORGANISM: Homo sapiens
 41 IMMEDIATE SOURCE:
 42 CLONE: COLLAGEN ALPHA 1 (111)
 43 US-09-500-811-21
 44
 45 Query Match 6.98; Score 628; 108 41; Length 1078;
 46 Post Local Similarity 25.08; Prod No. 4 40-24;
 47 Matches 428; Conservative 67; Mismatches 466; Indels 452; Gaps 71;
 48
 49 UY 259 UYAVADKIAAP RLDRESEVENKLLPVSFVASTLPPVDTGNSLNKAV 614
 1b 26 PFAFPPPPGPGGSGHSGSPSYOG ---PFGFVAGLSPPVYATGISTGFG 81
 UY 615 IVVSQSNSSSARKAIPVY-----SSDPPILKPP 448
 1b 62 KQSFSPRPRLDRLDLPFGKIPAGLPGFGMKHGFQKNGEKELTATKJCHN 140
 UY 349 DLSDFQ-----LPR-----ERSLDTLRLQMLFPDEKETLGAASG ---GPGNPLVLLG 697
 1b 141 ---LHGNAGVGMGRGAGRGKRLVLAAGAKRNDKARNSNDLGGPGPATAGPS 199
 UY 498 PGRKEP-GPLQAMMAVSSTGKGPRLVAVPFGGHLVVFSEEMVFSMSSSGT 456
 1b 200 PGAKGVYR---ASVSSNCAPIQR ---GFGVAGVAGAGGPG---PFGTNSGSG 247
 UY 457 TGPDLIDMTPLQAMIKLQGFYERKKRQVVVVVSTQIMNVHCHDQV--- 510
 1b 248 KQ-----LPGPAJ-----MINSEMERN-----VFNVANSNLSVSWITGV 262
 UY 511 ---VAGJPPVQMTPESEMAVACTEPEFSGT---NMPHSLPPKGAIPNPPNSQMK ----- 561
 1b 263 MKARKTPTACANQAPGKAGAGPGRKAKAGLEGGPGRERPAJLIVYAKKPEKDKNS 422
 UY 562 -----LPGPAJ-----MINSEMERN-----VFNVANSNLSVSWITGV 596

RESULT 5
 US-09-500-811-21
 Sequence 21: Application US-09-500-811
 Patient No. 642414
 GENERAL INFORMATION:
 APPLICANT: GYSLT, Inc
 APPLICANT: Bondy, Mattie

Db 555 P-----GSPGA--PGLTPGGLPG-----SPGAP--GPGGL 664
 QY 1084 LQPLSHSNOMSPDVAQVNTPE--PRAVPMGLQMSHNPIMHOSQEPVMEVPMPPGAP 1141
 Db 585 P-----GSPGA--PGLTPGGLPG-----SPGAP--GPGGL 664
 QY 1142 PTVOSPVQVFFPPLNDSQAGSPGAGMPPGGLDQKPSNLPVSNAAALN 1201
 Db 618 PAAAPTP-----GPGGLPGS--PAAAGTGG-----PAAAGTGG 664
 QY 1202 DSFTVANSMSVFTDPLQVTFHGAQGLHEDLSRIHSEKPSQGLVYPPG 1256
 Db 655 GPGGLPGS-----PAAAGTGG-----PAAAGTGG 679
 QY 1257 VPGKQPG-----PG--PQSHMOQMGQAPPMALPQM--GAPVAGTGLPGLTAS 1310
 Db 680 APGTPGGLPGSAGGATPGGLPGS-----PAAAGTGGGLVPSI--GAPG 728
 QY 1311 PGNPMPPAFLOQNMGPPIHKKHSPQOSTPQGLTMSNPAAVQMTPOKQSTACTY 1370
 Db 729 PG-----PQGLPG-----SPGAPTPPGGLGSPGA--PGLTPGGLPG 766
 QY 1371 HPGVAPSPGMMMSKQMGCPQOMLTPQMRPROMADVPMGP--SGP 1424
 Db 767 SPAPGTPG-----PGLTPGSG--APGTPGGLPGSPAPGTPGGLPGS 815
 RESULT 11
 US-08-642-255 101
 Sequence 101, Application US/08642255
 Patent No. 5776249
 GENERAL INFORMATION:
 APPLICANT: CARPILLO, Joseph
 APPLICANT: PERRAR, Franco A.
 TITLE OF INVENTION: High Molecular Weight Cellulose-Like
 NUMBER OF SHEETS: 145
 CURRENT INVENTOR ADDRESS:
 ADDRESS: FLEHER, HOMERICH, FESI, ALBERTON & BEMER;
 STREET: 4 Embarradero Center, Suite 4400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent to Release #1.0, Version #1.40
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/642,255
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: ROWLAND, William L.
 REGISTRATION NUMBER: 20,015
 REFERENCE/WORK NUMBER: A5506 4/BIR
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 494-8700
 TELEFAX: (415) 494-8771
 TELEEX: 910 277299 FBI OR
 INFORMATION FOR SEQ ID NO: 101:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 817 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-642-255-101
 Query Match 6.78; Score 515.5; Db 1: Local 6.67;
 Post Local Similarity 26.08; Pred. No. 10-23;
 Matches 311; Conservative 50; Mismatches 499; Indels 437; Gaps 74;

QY 275 DRESVPE NKLTPVSGVASSLTPPLPTGPNLSLNNAAV LVSQSSSSAAER 629
 Db 10 DQNPVPLQNLK--AAHVFASHPMAVATPGVGLPNSVATGTFVGLPNSVATG 67
 QY 430 APPP--PVNSHPTLGLNPGHLSQGLHFRKSLGTLRQMLPPEKTE PAVSSA 987
 Db 68 PGLQGLPNSVATGNG--PGLT-----PNSVATG--116 106
 QY 488 PGNPVP--LGGGKRPGLTQAMMASSSLGKGPRVAVAPPTGHLVYNS 442
 Db 101 PGLTPSPVAPGTPVGLPGSNMAVNTVGLTNSVGA--PGLT 300PNSVATG 156
 QY 443 DEKVPSSMSNSSTLIGRHLDRPTQGLAMLKQGFYEFERKKRGLVVVQGLSLNNV 502
 Db 157 GTPGGLVSVARG-----TP----- 774
 QY 503 DQHPGVVQVQTPPVQMTPEHMAVATPEPSGLNTRPSLTPROMAPPM--SNQML 562
 Db 175 ----GPGGLPGSAP--GTPGGLPG--SPGAPTPPGGLP--GSPVATGLPQG 224
 QY 563 PGPACMLNSMGPVNTN--ASRQGLSGVSMPPVKKTPQKNTVQGLFESQGP 619
 Db 225 PGSPG--APGTPGGLPNSHGAATPGVGLPNSVAPPTPGVGLGSPG 276
 QY 620 RFPNPGLSPEHFGQGLAKQGLTLPQMMMLRPSMKNNMLPNSGRKRPNNNTTP 679
 Db 277 GTPGGLPGS-----PG-----APGTPGGLPGS 694
 QY 680 TPVPSLSPGTPKGLPQPM--PQKLEPQVPSMGQVNLVNMSSNSQMT 745
 Db 405 ---ANGTPG--PGLTNSPAGTGG-----PGLTNS-- 882
 QY 746 KRRFAGVPEMLKLPQASMLPQOKVYLPENHVGVEGMDQDTLPSG 794
 Db 433 -----PQA-----PGLTPGGLPNSVATGLPGLPGS 862
 QY 796 SLRLNLPETPLQGRNSLSHPTPLNLSNLSLAP PGLTQKGLT 849
 Db 464 -----PAAVTPGGLPNSVATGP--GPGGLPNSVAPPTPGVGLPNSVATG 412
 QY 850 VANSQ--VHSVPLNLSKTPHVOVSPMLSPSNLKSPLLSQALMA 907
 Db 413 TPVGLTPNSVAPPTGP-----QGLTSGP--ANGTPG 918PNSVATG 401
 QY 908 PVLVNSAASVHLKSLAP--SPQWLSKPLQSGTPENKAPLMSVAM 962
 Db 462 QGLPNSVA-----PGLTPGGLPNSHGAATGP-----QGLTNSVATGLPQGP 410
 QY 963 GNVSGGTPPLASPAVSNLPSSLPYTPMPEPLSNQNLSSLMSSKSPVNS 1022
 Db 513 GSPGATGTPG--QGLPNSVAG--TPVGLPNSVATGTP-----GPGGL 564
 QY 1023 PLYHAKTIVASSDQSPFASPNLPSSNNMGMNTINUNTRSGNTPVMPPLSP 1062
 Db 555 P-----GSPGA--PGLTPGGLPG-----SPGAP--GPGGL 664
 QY 1084 LQPLSHSNOMSPDVAQVNTPE--PRAVPMGLQMSHNPIMHOSQEPVMEVPMPPGAP 1141
 Db 585 P-----GSPGA--PGLTPGGLPG-----SPGAP--GPGGL 664
 QY 1142 PTVOSPVQVFFPPLNDSQAGSPGAGMPPGGLDQKPSNLPVSNAAALN 1201
 Db 618 PAAAPTP-----GPGGLPGS--PAAAGTGG-----PAAAGTGG 664
 QY 1202 DSFTVANSMSVFTDPLQVTFHGAQGLHEDLSRIHSEKPSQGLVYPPG 1256
 Db 655 GPGGLPGS-----PAAAGTGG-----PAAAGTGG 679
 QY 1257 VPGKQPG-----PG--PQSHMOQMGQAPPMALPQM--GAPVAGTGLPGLTAS 1310
 Db 680 APGTPGGLPGSAGGATPGGLPGS-----PAAAGTGGGLVPSI--GAPG 728

